Signs of Colloquialization

Three Corpus-Based Case Studies

Mihail Iosef

Master's Thesis
submitted to
Faculty of Humanities

Department of Literature,
Area Studies
and
European Languages

UNIVERSITETET I OSLO

May, 2013

Signs of Colloquialization: Three Corpus-Based Case Studies

2013

Signs of Colloquialization: Three Corpus-Based Case Studies

Mihail Iosef

http://www.duo.uio.no/

Trykk: Reprosentralen, Universitetet i Oslo

Summary

The thesis examines, synchronically and diachronically, three ongoing linguistic changes in written English, in its two major varieties, British and American English. The three phenomena in question are: an increasing use of contracted forms (mainly, but not exclusively, verbal) observed in present-day written English; an increasing use of the word *like* with several colloquial functions, particularly with its quotative function, i.e. when *like* is used to introduce direct speech; and an increasing use of phrasal verbs with *up* and *out*. For example, the use of a non-standard contraction such as *gonna* has increased, since its first attestation back in the 1910s, by 189 times in written American English. Similarly, the informal use of the word *like* to introduce a quotation has increased by 68.5 times in written American English, since the early 1980s, the period when this use of *like* popped up in the English language. The use of phrasal verbs with *up* has increased by over 3 times over the last two centuries in American English. These are only a few examples of the kinds of phenomena the thesis deals with.

The analysis is approached from the perspective of what Mair (1997) termed 'the colloquialization of written English', i.e. the growing influence of speech on written language, reflected by an increased number of informal, colloquial and even non-standard linguistic features commonly associated with speech, in the written language.

The investigation has been carried out with massive corpus support from the largest English corpora of the moment: the Corpus of Contemporary American English (COCA), the Corpus of Historical American English (COHA), the British National Corpus (BNC), the Google Books American English Corpus[1], and the Google Books British English Corpus. The synchronic dimension of the thesis focuses on the quantitative analysis of the corpus data across the various genres in COCA and the BNC, while the diachronic side scrutinizes the evolution of the three linguistic phenomena over a period of time of roughly two centuries, since the 1810s up to the present, in COHA and the Google Books corpora. The results are more than obvious and generally support the colloquialization hypothesis put forward by Mair, although not entirely, as particular findings in the thesis will show.

[1] Not to be confused with the Google Books web service from Google Inc.

In loving memory of my parents, Constantin and Ana-Paulina Iosef

Preface

Although a master's thesis is, to a large extent, the result of individual work, it would however not be possible without the support of others. Thus, I would like to thank my supervisor, Gjertrud Flermoen Stenbrenden, for her constant and valuable advice and encouragement throughout the elaboration of the thesis. I would also like to thank Johan Elsness for inspiring, even if unintentionally, the research topic of the thesis, and for introducing me and my classmates to the universe of electronic corpora. Thanks to student advisor Tor Erik Johnsen for his administrative support during the entire period of study. Thanks to Tauqeer Ahmad Saadi for the exchange of general, but useful ideas about our theses. We have both simultaneously experienced the same 'pains' of creation of a master's thesis. Thank you, Silviu Mărgărit for taking care of my personal business back home, while I have been away to the beautiful Land of the Fjords, and thus, for indirectly helping me to complete this work. Last but not least, I would like to thank everyone at the University of Oslo, particularly at the Faculty of Humanities, for providing an enjoyable learning experience in a relaxed and modern environment. Since this thesis is about speech habits spreading into the written norm, I will conclude this preface in a style which signals the kind of issues which will be dealt with in the next chapters, and will tell everyone, mentioned and not mentioned here: it's been awesome to hang out with you guys.

Mihail Iosef

Oslo, May 2013

Contents

Summary V
Preface IX
1. Introduction and Background 1
1.1. General Remarks 1
1.2. Three Key Terms: Genre, Register, and Style 2
1.3. Colloquialization in the Literature 4
1.4. Aim and Scope 11
2. Method of Analysis 13
3. Contractions 16
3.1. Standard Contractions 16
3.2. Negative Contractions 21
3.3. Non-Standard Contractions 25
3.4. Other Contractions 37
3.5. Conclusions on Contractions 40
4. Colloquial *Like* 42
4.1. A Few Theoretical Considerations 42
4.2. Corpus Findings 46
4.2.1. Quotative *Like* 46
4.2.2. *Like* as Adverb 51
4.2.3. *Like* as Filler 53
4.2.4. *Like* as Subordinating Conjunction 57
4.3. Conclusions 59
5. Phrasal Verbs and Colloquialization 61
5.1. General Remarks 61
5.2. A Brief History of the English Verb-Particle Constructions 62
5.3. The Main Features of Present-Day Verb-Particle Constructions 63
5.4. *Up* 70
5.4.1. Values of *Up* in the Literature 70
5.4.1.1. Historical Values 70
5.4.1.2. Contemporary Values 74

5.4.2. Corpus Findings 83
5.4.2.1. Continuous VPCs with *up* 83
5.4.2.2. Discontinuous VPCs with *up* 96
5.4.3. Conclusions 112
5.5. *Out* 113
5.5.1. The Semantics of *out* 113
5.5.1.1. Historical values of *out* 114
5.5.1.2. Contemporary values of *out* 116
5.5.2. Corpus Findings 121
6. Conclusions 133
Appendix 135
Bibliography 137

Table 1. Normalized frequencies of verb contractions in the spoken and written material of COCA 16
Table 2. Normalized frequencies of verb contractions by genre in COCA 17
Table 3. Semantic classification of phrasal verbs (based on Thim 2012: 13) 64
Table 4. Combinational variants of monosylabic light verbs with various particles (based on Elenbaas 2007) 69
Table 5. Normalized frequencies of discontinuous VPCs with *up* in the written material of COCA and BNC, divided by genre 99
Table 6. Normalized frequencies of the discontinuous VPCs with *up* in the Magazine subgenres of COCA 103
Table 7. Normalized frequencies of the discontinuous VPCs with *up* in the Newspaper subgenres of COCA 104
Table 8. Normalized frequencies of the discontinuous VPCs with *up* in the Newspaper subgenres of BNC 105
Table 9. Normalized frequencies of the discontinuous VPCs with *up* in the Academic subgenres of COCA 106
Table 10. Normalized frequencies of the discontinuous VPCs with *up* in the Academic subgenres of BNC 107
Table 11. The search strings used in chapter 3 135
Table 12. The search strings used in chapter 4 135
Table 13. The search strings used in chapter 5 136

Figure 1. Diachronic view of contraction *'ll* in COHA 18
Figure 2. Diachronic view of contraction *I'm* in COHA 18
Figure 3. Diachronic view of contraction *it's* in COHA 19
Figure 4. Diachronic view of contraction *'re* in COHA 19
Figure 5. Diachronic view of contraction *'d* in COHA 20
Figure 6. Diachronic view of contraction *'ve* in COHA 20
Figure 7. Frequencies per million words of *cannot* in the spoken vs. written sections of COCA 21
Figure 8. Frequencies per million words of *can't* in the spoken vs. written sections of COCA 22
Figure 9. Frequencies per million words of *cannot* in the written material of COCA, divided by genre 22
Figure 10. Normalized frequencies of *can't* in the written material of COCA, divided by genre 23
Figure 11. Diachronic view of *can't* in COHA 23
Figure 12. Frequencies per million words of *n't* in the spoken vs. written sections of COCA 24
Figure 13. Normalized frequencies of *n't* in the written material of COCA, divided by genre 25
Figure 14. Diachronic view of *n't* in COHA 25
Figure 15. Frequencies per million words of *ain't* in COCA and BNC in spoken vs. written material 26
Figure 16. Frequencies per million words of *ain't* in the written material of COCA and BNC, divided by genre 27
Figure 17. Diachronic view of *ain't* in COHA 27
Figure 18. Frequencies per million words of *dunno* in COCA and BNC in spoken vs. written material 28
Figure 19. Frequencies per million words of *dunno* in the written material of COCA and BNC, divided by genre 29
Figure 20. Diachronic view of *dunno* in COHA 29
Figure 21. Frequencies per million words of *gimme* in COCA and BNC in spoken vs. written material 30
Figure 22. Frequencies per million words of *gimme* in the written material of COCA and BNC, divided by genre 31
Figure 23. Diachronic view of *gimme* in COHA 31
Figure 24. Frequencies per million words of *gonna* in COCA and BNC in spoken vs. written material 32
Figure 25. Frequencies per million words of *gonna* in the written material of COCA and BNC, divided by genre 33
Figure 26. Diachronic view of *gonna* in COHA 33
Figure 27. Frequencies per million words of *gotta* in COCA and BNC in spoken vs. written material 34
Figure 28. Frequencies per million words of *gotta* in the written material of COCA and BNC, divided by genre 35
Figure 29. Diachronic view of *gotta* in COHA 35

Figure 30. Frequencies per million words of *wanna* in COCA and BNC in spoken vs. written material 36
Figure 31. Frequencies per million words of *wanna* in the written material of COCA and BNC, divided by genre 36
Figure 32. Diachronic view of *wanna* in COHA 37
Figure 33. Frequencies per million words of *lotta* in COCA and BNC in spoken vs. written material 38
Figure 34. Frequencies per million words of *lotta* in the written material of COCA and BNC, divided by genre 38
Figure 35. Diachronic view of *lotta* in COHA 39
Figure 36. Diachronic view of *lotta* in Google Books: British English 39
Figure 37. Diachronic view of *lotta* in Google Books: American English (Tokens) 40
Figure 38. Frequencies per million words of quotative *like* in COCA and BNC in spoken vs. written material 47
Figure 39. Normalized frequencies of quotative *like* in the written material of COCA and BNC, divided by genre 47
Figure 40. Diachronic view of quotative *like* in COHA 49
Figure 41. Diachronic view of quotative *like* in Google Books American English 49
Figure 42. Diachronic view of quotative *like* in Google Books British English 50
Figure 43. Diachronic view of quotative *like* in COCA (1990 – 2012) 50
Figure 44. Normalized frequencies of *like* as colloquial adverb in COCA and BNC in spoken vs. written material 52
Figure 45. Normalized frequencies of *like* as colloquial adverb in the written material of COCA and BNC, divided by genre 52
Figure 46. Diachronic view of *like* as colloquial adverb in COHA 53
Figure 47. Normalized frequencies of *like* as a filler in COCA and BNC in spoken vs. written material 54
Figure 48. Normalized frequencies of *like* as a filler in the written material of COCA and BNC, divided by genre 55
Figure 49. Diachronic view of *like* as a filler in COHA 55
Figure 50. Diachronic view of *like* as a filler in COCA (1990 – 2012) 56
Figure 51. Normalized frequencies of the filler *like* placed at the end of the utterance in the written material of COCA and BNC, divided by genre 56
Figure 52. Normalized frequencies of *like* as subordinating conjunction in COCA and BNC in spoken vs. written material 57
Figure 53. Normalized frequencies of *like* as subordinating conjunction in the written material of COCA and BNC, divided by genre 58
Figure 54. Diachronic view of *like* as subordinating conjunction *in* COHA 58
Figure 55. Diachronic view of *like* as subordinating conjunction in COCA (1990 – 2012) 59
Figure 56. Frequencies per million words of the continuous VPCs with *up* in COCA and BNC, in the spoken vs. written material 83
Figure 57. Normalized frequencies of the continuous VPCs with *up* in the written material of COCA and BNC, divided by genre 84

Figure 58. Distribution of the continuous VPCs with *up* in the subgenres of Fiction in COCA 86
Figure 59. Distribution of the continuous VPCs with *up* in the subgenres of Fiction in BNC 86
Figure 60. Distribution of the continuous VPCs with *up* in the subcategories of Magazine in COCA 88
Figure 61. Distribution of the continuous VPCs with *up* in the Newspaper subcategories of COCA 89
Figure 62. Distribution of the continuous VPCs with *up* in the Newspaper subcategories of BNC 91
Figure 63. Distribution of the continuous VPCs with *up* in the Academic subcategories of COCA 91
Figure 64. Distribution of the continuous VPCs with *up* in the Academic subcategories of BNC 92
Figure 65. Diachronic view of the continuous VPCs with *up* in COHA 94
Figure 66. Diachronic view of the continuous VPCs with *up* in COCA (1990 – 2012) 94
Figure 67. Diachronic view of the continuous VPCs with *up* in Google Books American English 95
Figure 68. Evolution of the continuous VPCs with *up* in COHA by genre 95
Figure 69. Diachronic view of the continuous VPCs with *up* in Google Books British English 96
Figure 70. Frequencies per million words of discontinuous VPCs with *up* in COCA and BNC 98
Figure 71. Normalized frequencies of the discontinuous VPCs with *up* in the Fiction subgenres of COCA 101
Figure 72. Normalized frequencies of the discontinuous VPCs with *up* in the Fiction subgenres of BNC 101
Figure 73. Diachronic view of the discontinuous VPCs with *up* in COHA 107
Figure 74. Evolution of the discontinuous VPCs with *up* in COHA by DO type 108
Figure 75. Evolution of the discontinuous VPCs with *up* in COHA by genre 109
Figure 76. Diachronic view of the discontinuous VPCs with *up* in COCA (1990 – 2012) ... 110
Figure 77. Diachronic view of the discontinuous VPCs with *up* in Google Books American English 111
Figure 78. Diachronic view of the discontinuous VPCs with *up* in Google Books British English 112
Figure 79. Frequencies per million words of the VPCs with *out* in COCA and BNC, in the spoken vs. written material 122
Figure 80. Normalized frequencies of the VPCs with *out* in the written material of COCA and BNC, divided by genre 123
Figure 81. Distribution of the VPCs with *out* in the subgenres of Fiction in COCA 123
Figure 82. Distribution of the VPCs with *out* in the subgenres of Fiction in BNC 124
Figure 83. Distribution of the VPCs with *out* in the subcategories of Magazine in COCA .. 125
Figure 84. Distribution of the VPCs with *out* in the Newspaper subcategories of COCA 125
Figure 85. Distribution of the VPCs with *out* in the Newspaper subcategories of BNC 126

Figure 86. Distribution of the VPCs with *out* in the Academic subcategories of COCA 127
Figure 87. Distribution of the VPCs with *out* in the Academic subcategories of BNC 128
Figure 88. Diachronic view of the VPCs with *out* in COHA 128
Figure 89. Diachronic view of the VPCs with *out* in COCA (1990 – 2012) 129
Figure 90. Evolution of the VPCs with *out* in COHA by genre 129
Figure 91. Diachronic view of the VPCs with *out* in Google Books American English 130
Figure 92. Diachronic view of the VPCs with *out* in Google Books British English 131
Figure 93. *Freak out* in COHA and Google Books Corpora (British and American English; normalized frequencies) 132
Figure 94. *Screw up* in COHA and Google Books Corpora (British and American English; normalized frequencies) 132

1. Introduction and Background

1.1. General Remarks

The tendency, in English in general, and in the English spoken in the United States in particular, to simplify language and to reduce complex concepts, sometimes even to mere abbreviations, is by no means a new fact. Mencken (1947: 92) notes that in America, such processes have been noticeable since colonial times, when the first colonists had to coin new words to describe the brand-new, unfamiliar environment they were confronted with (cf. Mencken 1947: 3). One of the most common Americanisms, for instance, *O.K.*, also spelt *OK*, *o.k.*, or *okay*, dates back not exactly from the colonial age, but rather from the early times of the Republic, more precisely from 1839, and is an abbreviation of a popular slang term at that time, *oll korrect* which, in turn, is a jocular misspelling of *all correct* (cf. *Collins Online Dictionary*). Mencken (1947: 93) gives examples of early language change processes, such as nouns turned into verbs by the simple addition of the preposition *to* before the noun, e.g. *to engineer.* A more recent example of similar change can be the same *O.K.* from above, which joined the category of a verb, meaning 'to approve', in ca. 1988 (cf. *The Concise New Partridge Dictionary of Slang and Unconventional English* 2008: 469). And the list could go on with the latest coinages of the now in vogue verbs *to google, to skype* and *to wikipedia*, derived from the corresponding proper nouns. But this is not to say that the English spoken in the British Isles is not innovative at all, despite its notorious conservatism. However, Mencken holds that American English is less prone to the 'suffocating formalization' that Standard British English exhibits:

> That is to say, we incline toward a directness of statement which, at its greatest, lacks restraint and urbanity altogether, and toward a hospitality which often admits novelties for the mere sake of their novelty, and is quite uncritical of the difference between a genuine improvement in succinctness and clarity, and mere extravagant raciness. (Mencken 1947: 94)

Mencken maintains that, like any other language, English changes, innovates, and embraces vulgarisms and informality because these elements meet the practical communication needs of the language users; they do so mainly in speech, but sometimes in the written norm, too.

> The history of English, like the history of American and of every other living tongue, is a history of vulgarisms that, by their accurate meeting of real needs, have forced their way into sound usage, and even into the lifeless catalogues of the grammarians. (Mencken 1947: 96)

Although Mencken does not use the term 'colloquialization' for obvious reasons (the term simply did not exist at that time), his point is that a tendency from formal to informal, from the rigid restraints of Standard English to a more relaxed, 'liberal' and innovative form of linguistic communication, has been evident in American English even since its earliest times, and this can be identified with what today is called colloquialization.

1.2. Three Key Terms: Genre, Register, and Style

Before moving on to a more in-depth discussion on colloquialization and its effects on language change, I find it useful to briefly review three terms which, although they do not constitute the object of study of this thesis themselves, are basic for its purpose. And these terms are **genre, register,** and **style**.

Genre in this thesis does not refer to literary categories, such as tragedy, comedy, novel, or short story, but, as the *Longman Dictionary of Language Teaching and Applied Linguistics* defines it, to:

> A type of discourse that occurs in a particular setting, that has distinctive and recognizable patterns and norms of organization and structure, and that has particular and distinctive communicative functions. For example: business reports, news broadcasts, speeches, letters, advertisements, etc. In constructing texts, the writer must employ certain features conventionally associated with texts from the genre in which he or she is writing. In reading a text the reader similarly anticipates certain features of the text based on genre expectations. (*Longman Dictionary of Language Teaching and Applied Linguistics* 2010: 245)

Register and **style**, on the other hand, might seem, at first glance, confusing, and while some linguists use the two notions interchangeably, others differentiate between them. Biber & Conrad (2009: 2) explain that, while the linguistic features of a particular register depend on the situational context of the text and are always functionally motivated, style features are not functional, but they rather reflect the artistic or aesthetic preferences of a certain writer or

historical age. Biber & Conrad point out that registers can be spoken or written, and that there are several fundamental differences between the two. One such difference is the interpersonal function of the spoken registers; that is, at the most basic level, the primary goal of speakers in a communicative act is to express their feelings and attitudes, rather than to communicate some piece of information (cf. Biber & Conrad 2009: 85). Conversely, written registers are primarily concerned with communicating information, or with explaining and describing facts, rather than with developing a personal relationship between the writer and the reader. Moreover, spoken registers are usually spontaneous and interactive, which leaves little, and sometimes no, time for planning, editing and revision. In contrast, written registers allow writers plenty of time to conceive, to draft, and if needed, to subsequently alter the language of their texts. However, not all spoken registers are interpersonal, just as not all written registers are entirely descriptive or informative. A public speech is not interpersonal, though quite often it is intentionally designed to sound like that. On the other hand, some written registers, such as novels or newspaper articles, often borrow features of spoken registers in the form of quotations, or other narrative devices in order to simulate oral communication. At the same time, a letter or an e-mail can be interpersonal as well, focusing on the development of a personal relationship between the addresser and the addressee. The most basic of all spoken registers, conversation, has a number of characteristic features which are not usually found in the written registers, unless the above mentioned speech simulation is employed. Among the most common features which differentiate spoken from written registers, Biber & Conrad mention the frequent use of contractions and other structural reductions, such as *watcha* for *what are you*, or *[I] can't help you* (the *I* between the square brackets is sometimes omitted in conversation), repetitions, which usually occur when the speaker does not know what to say, "attention getters", e.g. *hey*, minimal responses, e.g. *hmm?, okay*, or semi-modals such as *going to, have to, got to, had better* and their reduced forms *gonna, hafta, gotta,* and *better*, to name only a few of them (cf. Biber & Conrad 2009: 90).

Although written registers share a number of common features, such as the general informative purpose and the much longer available time for planning, editing, and revising the language, they however differ in many respects, and each register has a particular set of features which help identify it. For instance, one notable difference between newspaper writing and academic prose is that while the former focuses on reporting important current events, the latter is expected to analyze and explain, not just report (cf. Biber & Conrad 2009: 118). Fiction, on the other hand, is one of the most complex registers, due to the wide range of

stylistic choices a writer can make, choices which have immediate linguistic implications on the text. For instance, if the writer chooses to narrate the story from a first person perspective, a high frequency of the first person pronoun *I* is obviously expected to occur in the text. At the same time, in a first person narrative, perceptions, thoughts and attitudes of the narrator are usually reported, and as a consequence, a large number of *that* and *to* complement clauses usually occur in such a text, where the verb in the main clause expresses a personal standpoint about the information comprised in the complement clause (cf. Biber & Conrad 2009: 133). If, on the contrary, the story is told from the perspective of an omniscient third person narrator, the number of first person personal pronouns is expected to decrease, and a large number of mental verbs which subordinate complement clauses are usually employed in the narration (cf. Biber & Conrad 2009: 134). If the author decides to report a great deal of dialogue of the characters, then the fictional style comes to resemble real-time, face-to-face conversation, and frequent questions, contractions, ellipses, present tense verbs, and second person pronouns (in addition to first person pronouns) will occur in a text that adopts this style of fiction (cf. Biber & Conrad 2009: 135). In the following chapters I will try to show, using corpus evidence, how such linguistic features typical of the spoken language have spread into the various written genres of the two major varieties of English, British and American English.

1.3. Colloquialization in the Literature

Since its coinage by Mair in 1997, the term "colloquialization" has long been debated in works dealing with language change in the 20th and 21st centuries. According to Mair, the term defines a significant stylistic move, away from an elaborated written norm which is positioned far from the informal spoken style, towards a written norm which is closer to the speech habits; a move away from a norm characterized by formality towards a norm marked by informality (cf. Mair 2006: 187). Or, as he and Hundt put it earlier in a simpler way, colloquialization could be defined as "a narrowing of the gap between spoken and written norms" (Hundt & Mair 1999: 221).

According to Mair (2006: 183), the default instance of linguistic communication is informal, face-to-face conversation. Compared to this, other types of communication, such as polite, informal speech or written language require, for functional purposes, some degree of structural elaboration. Very often, this elaboration is redundant, conferring the written

language "an element of arbitrariness, artifice, ritual, and fashion" (Mair 2006: 183). Among the features typically associated with the informal style, Mair mentions sentence-initial conjunctions, first and second person references, and emphatic particles such as *really* and *just* (cf. Mair 2006: 183). The formal register, on the other hand, is characterized by such features as nominalizations and a frequent use of the passive voice, Mair notes.

Many linguists have used corpora to find evidence of the already noticeable language changes resulted from colloquialization. Smith (2002) carried out an investigation on the spread of the progressive in recent British English, using the LOB-FLOB pair of corpora, and found that the progressive forms are indeed spreading, especially in the present tense, one of the possible reasons for such a trend being the colloquialization of the written language under the influence of informal speech (cf. Smith 2002: 327). Mair (1997) concluded, after his extensive corpus-based investigation on the *be going to* + infinitive construction, that this is spreading in written English, especially in its American variety, at the expense of the *will/shall*-future, as a result of the colloquialization perceptible over the last three decades (cf. Mair 1997: 1541). Smith (2002: 327) cites Hundt and Mair (1997) who, in turn, have pointed out that the growing use of phrasal verbs with *up* in writing, and the decline in the use of the archaic preposition *upon* are further evidence of colloquialization. Mair (2006: 189) argues that an increased frequency of verb contractions and negative contractions in writing can be sensed in the press sections of the LOB, FLOB, Brown and Frown corpora. Mair claims that the reason for such a growing trend is that the authors wish their texts to be more appealing and accessible to the readers.

Leech *et al.* (2009: 106) note various changes occurring in English over a time span of thirty years, since 1961 to 1991, the period covered by the Brown Frown and LOB F-LOB pairs of corpora. One such change observed by Leech *et al.* is the high usage rate of semi-modals, such as *be going to, have got to, want to,* and their non-standard contracted forms *gonna, gotta* and *wanna* in the spoken material of the BNC, as opposed to the very low rates of the same phonetic reductions in the written sections of the Brown family of corpora. Moreover, they notice further phonetic reduction, in addition to the contraction itself, namely the ellipsis of a whole word, usually the auxiliary; for instance, the use of *gonna* alone, instead of BE *gonna*, as in the example below taken from Leech *et al.* (2009: 106):

(1) So what is it, Lieutenant, you *gonna* read me my rights? (Frown L13)

Another example of such auxiliary ellipsis is the frequent omission of *had* in *had better,* as in the following excerpt taken from Leech *et al.* (2009: 107):

(2) You shake your hands with that guy and you *better* count your fingers when you walk away. (Frown N10)

The elision of *d* in the BE *supposed to* construction is a further instance of common phonetic reduction observed by Leech *et al.* (2009: 107). The *get*-passive, as an informal alternative to the *be*-passive, e.g. *the car got sold* vs. *the car was sold,* seems to have also increased, according to Leech *et al.* (2009: 146). On the other hand, the *be*-passive seems to have dramatically declined in academic prose, while a marked increase in the use of first person personal pronouns can be perceived in the same genre (cf. Leech *et al.* 2009: 152). However, Leech *et al.* note that *be*-passives are not necessarily restricted to the formal style, but can be employed in formal speech or writing as well, as in the following example taken from Leech *et al.* (2009: 153), where the verb *nick* is used with its slang meaning 'to take into legal custody, to arrest':

(3) […] racially harassing Golliwog, the woman policeperson told him. Golly! Noddy exclaimed. Right. You're *nicked* as well. (F-LOB R05).

For all these changes, Leech *et al.* propose a number of explanatory processes, among which are **grammaticalization, colloquialization**, and changes resulting from contact between various English dialects, the most notable such process being known as '**Americanization**' (cf. Leech *et al.* 2009: 237). As far as grammaticalization is concerned, Leech *et al.* note that in American English, for instance, the use of constructions such as *have to, have got to,* and *need to* have overrun the core modal *must* (cf. Leech *et al.* 2009: 238). They also underline the key role of style and register in the spread of language innovation; that is, some genres are more open to embracing newly grammaticalized forms, while some others are more resistant to such changes, and they point out that these forms come to be accepted faster in the spoken than in the written language (cf. Leech *et al.* 2009: 239). While grammaticalization is a language-internal process, colloquialization is rather a discourse-based phenomenon; that is, while the result of grammaticalization is the addition of new forms to the language system, colloquialization deals with shifting preferences among various already existing options (cf. Leech *et al.* 2009: 237). Leech *et al.* note, as I shall discuss in more detail in 5.4.2.1, that the journalese of the last thirty years has seen a marked move away from the traditional formal

style to a seemingly deliberately simulated direct writer-reader interaction intended to mimic oral communication (cf. Leech *et al.* 2009: 239). Similarly, fiction has seen, according to Leech *et al.*, a turn to registers similar to the spoken language since the 19th century. Leech *et al.* (2009: 240) claim that the use of contractions has dramatically increased over the last forty years, as evidenced by their research in the Brown, Frown, LOB, and F-LOB corpora. Particularly the first person plural imperative contraction *let's* is, according to Leech *et al.*, a clear illustration of colloquialization, as this has increased inversely proportional to the decrease of the full form *let us*, especially in American English (cf. Leech *et al.* 2009: 240). As the most speech-like written genre, fiction has the highest usage rate of *let's*, while the full form *let us* is almost completely absent from American English fiction, according to Leech *et al.* (2009: 240). Another sign of colloquialization suggested by Leech *et al.* (2009: 242) is the increased use of questions in writing, since these are more common to speech. Again, the highest growth is found in fiction. In terms of punctuation, Leech *et al.* note that, beside the increased number of question marks, a similar growth in the use of full stops may also be considered evidence of colloquialization, as more full stops in a text mean more but shorter and simpler sentences, while, conversely, fewer full stops would indicate longer sentences, and therefore, more complex syntax. And the data from the four corpora revealed an increased number of full stops in American English, but curiously, a decrease in British English, which might indicate a sort of anti-colloquialization trend in this variety (cf. Leech *et al.* 2009: 245). The significant increase in snippets of direct speech and quotations in both fiction and non-fiction is, according to Leech *et al.*, further evidence of colloquialization, and this extensive presence of quotation marks in texts of all genres requires a slight revision of the definition of colloquialization, as being not only "writing becoming more like speech but writing **containing** more speech" (Leech *et al.* 2009: 249; my emphasis). Expanding this, Mair (1998: 153) explains that colloquialization works on two levels. On a narrower, grammatical level, it manifests itself through an increased use of constructions traditionally associated with speech, such as the *be going to* future, the progressive, contractions, and phrasal verbs, to name only a few. On the broader level of textual organization, Mair claims, colloquialization consists of an increased use of direct speech and other stylistic features which suggest the oral mode in written language. It is the direct speech snippets in a text that contain the bulk of informal constructions common to the spoken language. The more direct speech fragments there are in a text, the more colloquial, informal and spoken features the text will contain. Furthermore, Mair differentiates between what he calls 'fast' and 'slow' genres, that is, between genres

which are more open to colloquialization, and genres which resist it, and he notes that the contrast between the two categories is even more obvious at the level of textual organization than at the level of grammar (cf. Mair 1998: 153). Mair argues that the prototypically 'slow' genre, characterized by rigid, formal conventions, is academic writing, while at the opposite pole are newspapers, which, according to Mair, have changed substantially over the last four decades, from a traditionally formal style to a much more relaxed, oral, and reader-friendly style (cf. Mair 1998: 154). Mair argues, as I shall discuss in more detail toward the end of this subsection, that the colloquialization of written English is not a mere linguistic phenomenon, but it is closely related to, and influenced by, the general socio-cultural context in the English speaking world (cf. Mair 1998: 154).

As far as Americanization is concerned, Leech *et al.* (2009: 253) argue that this fits a particular change pattern in which two or more language varieties undergo the same change process, but one of them takes the lead, while the others follow it. This means that Americanization is not necessarily the result of direct American influence on British English or other regional varieties through dialect contact, but it could simply indicate that a particular change common to American and British English (and possibly to other regional varieties, too) is more advanced in the American variety (cf. Leech *et al.* 2009: 254). On the other hand, Leech *et al.* point out that there are situations in which a particular change has opposite evolutions in the two main varieties of English, declining in one variety and increasing in the other. However, Americanization can be said to be closely interrelated with colloquialization, in that some processes by which speech habits spread into writing, and which originated in American English, are subsequently borrowed by British English, too. A good example in this respect is quotative *like*, whose use in the two main varieties of English I will analyze in much closer detail in a separate chapter. Not only is this linguistic phenomenon spreading in the variety in which it originated, American English, and in the other major variety of English, British English, but it is also spreading in several other regional varieties, such as Australian English, Canadian English, Newfoundland English, Scottish English, and Jamaican English (cf. Leech *et al.* 2009: 258). Therefore, the spread of quotative *like* can be regarded as both an Americanization and colloquialization process. However, Mair (2006: 183) argues that Americanization should not be regarded as a mere spread of specific phonetic, lexical, or grammatical features from American English into other regional varieties, but it should be viewed in a broader context, also including genre, style and discourse conventions, which are now dominant all over the English-speaking world, and even beyond.

Coming back to colloquialization, Westin (2002: 62) argues that the increased use of informal and conversational features, such as present tense verbs, NOT-negations, contractions, questions, and imperatives in her self-made Corpus of English Newspaper Editorials (CENE), comprising editorials from three large-circulation British newspapers (the *Daily Telegraph*, the *Guardian*, and *The Times*) and spanning the period 1900 – 1993, reflects a tendency towards more explicit and reader-friendly texts. According to Westin, this tendency is reinforced by a concomitant decrease in the use of features marking vagueness and implicitness, such as adverbial amplifiers, e.g. *absolutely, enormously, perfectly*, etc., and the pronoun *it*. On the other hand, contrary to Leech *et al.* (2009: 245), Westin's research reveals a tendency towards shorter sentences in the editorials of the three British newspapers analyzed, as a result of the reduced use of subordinate and relative clauses (cf. Westin 2002: 84). The reason for this trend might be, according to Westin, the same interest among authors in making the language of their editorials more reader-friendly, as shorter sentences with fewer subordinate clauses are easier to grasp than longer and more complex ones with a great deal of subordination involved. Westin concludes that, since the language used in newspapers reflects language use in society as a whole, the overall linguistic changes towards a more informal style observed in the English newspaper editorials support the assumption of a colloquialization of the written language (cf. Westin 2002: 165). But the increasing use of informal and speech-like features in writing is not only observable in present-day English. It seems to have been evident in late Modern English, too. To support his claims of early colloquialization, Smitterberg (2012) investigated NOT-contractions in the Corpus of Nineteenth-Century English (CONCE), and the data revealed significant growth rates in the non-expository genres of the corpus, such as Drama and Fiction, but a total absence from the expository genres, such as transcriptions of debates from the Houses of Parliament, and historical and scientific monographs. The only non-expository genre which showed neither a growth nor a decrease, but only a stable low usage rate, is Trials, consisting of transcriptions of 19th century trial proceedings (mainly in dialogue format). However, it is unclear whether this low rate in Trials, as well as the total absence from Debates, is a result of the fact that NOT-contractions were not used in speech in such contexts, or whether they were printed as full forms in the transcriptions (cf. Smitterberg 2012: 200). In addition to NOT-contraction, Smitterberg argues that a large number of other linguistic features, among which an increasing use of phrasal verbs and of the progressive, the spread of the BE *going to* construction as a way of expressing future tense, at the expense of *will* and especially of *shall*,

the use of *can* to indicate permission, and the rise of quantifiers such as *a lot of* at the expense of *much* and *many*, point to an early colloquialization of English in the 19th century (cf. Smitterberg 2012: 202).

Finally, it is worth mentioning here a number of extralinguistic factors which could account for the colloquialization hypothesis. Firstly, Biber & Finegan (1989) note that, by the 19th century, a general trend towards a popular literacy, put into practice through mass education, could be felt in both the British society, and in the New World.

> This trend was also reinforced by an overt interest in nature and a philosophical preference for naturalness and utility, which found expression in Romantic art, music, and literature. In prose, these Romantic attitudes resulted in a preference for an individual, colloquial self-expression rather than an elaborated, impersonal, and abstract style of argumentation (Biber & Finegan 1989: 514).

Biber & Finegan point out that "for many writers of the Romantic period, natural prose meant a colloquial style, reflecting conversation's place as the most basic mode of communication" (Biber & Finegan 1989: 514). These trends towards more colloquial, speech-like writing styles continued into the modern period:

> The development of a popular literacy fostered a shift towards more oral styles, reflecting the general levels of literacy competence. This shift was reinforced by an aesthetic preference for natural and colloquial styles. These attitudes continue to the present time, as reflected by college handbooks on writing, which praise direct, 'active' styles and warn against passive, nominalized, and structurally complex styles (Biber & Finegan 1989: 514).

Secondly, closely related to colloquialization is what Leech *et al.* (2009: 259) call the **democratization** of discourse conventions, i.e. a tendency "to phase out markers of distance, respect, superiority or inferiority, and to aim at the expression of greater equality and familiarity" (Leech *et al.* 2009: 259). A good example of linguistic democratization is the decline in the use of titular nouns, such as *Mr.*, *Mrs.*, and *Miss* as a prefix to names (cf. Leech *et al.* 2009: 259). Leech *et al.* point out that the given name alone, or followed by the surname, has become the new normal way of addressing or referring to a person.

Beal (2004: 6) notes that in Britain, the Industrial Revolution caused a massive migration of manpower from the traditionally agricultural areas to the newly emerging industrial centers, which in its turn, led to the occurrence of an urban working class, with its own urban dialects acting as a counterbalance to the influence of Received Pronunciation in speech, and Standard English in writing. In music halls, the entertainers of the moment could be heard singing songs written in these urban dialects, "some of which were to become anthems of local and class identity" (Beal 2004: 6). One such example is *The Blaydon Races*, which, according to Beal, "was sung by soldiers from the north-east of England in the trenches of World War I, and is still heard today from fans of Newcastle United Football Club" (Beal 2004: 6). The introduction, in 1840, of the so-called Penny Post in Britain, a cheap and uniform postal service, made it possible for more people than ever before to write and send letters, and consequently, through their inherently informal nature, personal letters may have contributed to the spread of speech features into writing (cf. Smitterberg 2012: 203).

Mair (2006), too, claims that the social context directly shapes language through a contextualized discourse, and changes such as the ones discussed so far point to a more informal and more 'oral' written English (cf. Mair 2006: 182). Mair argues that an increased social mobility, as opposed to the former relatively stable class-based hierarchies, and the tendency towards a more egalitarian society in the second half of the twentieth century, were, among others, reflected in a slip of popular written registers, such as letters, essays, and fiction towards the spoken registers (cf. Mair 2006: 185).

1.4. Aim and Scope

Assuming that the colloquialization hypothesis is true, the main questions I will try to answer in this thesis are: Is there an increase in the frequency of contracted forms and phrasal verbs with *up* and *out* in written English which can support the colloquialization theory? In addition, is there an increase in the informal uses of the word *like* in written English which can be considered evidence of colloquialization? If so, what is the current status and distribution of these three linguistic features in the two major varieties of English, British and American English? The expected answer, based on the previous studies, reviewed in the present introductory chapter, as well as on my own introspective knowledge of English, is that the three features intended to be investigated are on the rise in written English, as a consequence of colloquialization. To make predictions on the future evolution of language is

beyond the aim and scope of this thesis, as such predictions entail stepping on unsafe ground, but on the basis of the previously observed facts, we may assume that some of the terms considered non-standard by grammarians today, will, in the closer or more distant future, make their way into the written norm of language, while others will have only a meteoric existence, and will fade away, eventually falling into oblivion after their short-lived being. As an example, some of the verbs fashioned by the process of back-formation, described by Mencken (1947: 192) as being 'still on probation' at the time when he wrote the first edition of his book, that is, on the edge between standard and non-standard, neither accepted nor rejected by the scholars of the time, such as *to housekeep* or *to innovate,* are now established forms of Standard English. Similarly, since it is the language usage that dictates the norm, and not the other way around, we may assume that it is probably only a matter of time before such items as *to wikipedia* will have their own entries in the future editions of dictionaries, exactly as it has already happened with *to google* and *to skype*[2].

[2] The *Oxford English Dictionary* (Second Edition on CD-ROM, version 4.0) has a draft entry for *Google* as both intransitive and transitive verb, dated March 2006, but no entry for *Skype* and *Wikipedia,* while Collins Online Dictionary has an approved definition of the verb *to skype,* added on 31 October 2012, as well as for *to google,* but no entry for *to wikipedia* yet.

2. Method of Analysis

This study is primarily a quantitative one. Its aim is to answer questions like *what, where* and *when,* and to a lesser extent *why* and *how.* The corpus investigation has a contrastive, synchronic dimension, in that it compares data obtained from two different varieties of the same language, namely present-day English, and a diachronic dimension, as it tries to give an overview of the trends over a given period of time, from approximately 1810 to the present. The two varieties of English that I chose to investigate are British and American English. This choice is motivated by their current influential status, especially by the growing importance of the American variety as the new universal language of business, science and technology. Accordingly, the corpora that I chose to use are: the Corpus of Contemporary American English (COCA) and Google Books American English for the American variety of English, and the British National Corpus (BNC) and Google Books British English for the British variety. The choice of these particular corpora is motivated by several factors: first, their considerable size offers a massive amount of data to work with, like no other corpus does. The Google Books[3] family of corpora, in particular, comprises a stunning figure of more than 200 billion words to search through, and consists of four distinct corpora: Google Books American English (155 billion words), Google Books British English (34 billion words), Google Books Spanish (45 billion words), and Google Books One Million Books (89 billion words). Of the four corpora, I used two: Google Books American English for the queries concerning the American variety of English, and Google Books British English for the British variant.

Secondly, in the case of COCA, I opted for a 'monitor corpus', which continually expands over time, thus achieving a good degree of representativeness of the latest trends in language usage (cf. McEnery *et al.* 2012: 6). The other available option would have been a so-called 'balance-' or 'sample-corpus', such as the Brown family of corpora, which are structured according to a specific sample frame, and offer a 'snapshot'[4] of language over a certain period of time (cf. McEnery *et al.* 2012: 6). Since the members of the Brown family are no longer

[3] I am not talking here about the Google Books service offered by Google Company, but about the corpora created by the same tireless corpus enthusiast Mark Davies of Brigham Young University who also created COCA and COHA. Davies's Google Books corpora use the same data as Google Books by Google Inc., but offer a wider and more complex range of searches and features.

[4] Cf. McEnery *et al.* (2012: 9)

developed, they are representative of the language used until the early 1990s, and in this respect, they are 'frozen' in time. Moreover, they do not contain any spoken material – a shortcoming which cannot be overlooked, considering the purposes of this thesis. COCA, on the other hand, is a 'living' corpus and a combination of a sample and a monitor corpus, as it is structured according to a specific sampling frame (cf. McEnery *et al.* 2012: 7). Last but not least, I chose COCA for its ease of access and use, and I should also mention here that I accessed the BNC from the COCA's user-friendly interface. But the problem with using two or several corpora for comparative purposes is the difference between their structures. Of course, working with frequencies per million words (also termed normalized frequencies in this thesis) would solve the size difference, but what if the two corpora have a different categorization of genres, as is the case with the BNC and COCA? When accessing the BNC from the COCA interface, the extra genres in the BNC are grouped under the somewhat ambiguous categories 'miscellaneous' and 'non-academic'. These two extra categories are not found in COCA, and this poses problems of equivalence when comparing frequencies in terms of genres. To overcome this shortcoming, and for the sake of symmetry, I opted for comparing only the frequencies in the four genres common to both corpora, namely fiction, magazine, newspaper, and academic, ignoring the miscellaneous and non-academic categories. This was applied only in the case of the written material of the two corpora. Otherwise, when comparing the BNC written material as a whole, either to the spoken section of the BNC, or to the written material of COCA, the two extra categories were not ignored. To get the frequencies of the written material as a whole for a particular search, I added up the normalized frequencies of the written categories, and divided the total by the number of categories. For example, to get the overall written frequency of the contraction *'ll* in COCA, I used the following formula: 1,223.37 (the normalized frequency of *'ll* in 'fiction') + 579.81 (the normalized frequency of *'ll* in 'magazine') + 381.32 (the normalized frequency of *'ll* in 'newspaper') + 45.52 (the normalized frequency of *'ll* in "academic') divided by 4 (the number of written genres) = 557.50 (the overall normalized frequency in 'written') (see Table 1 and Table 2 for more details). In the case of the BNC, I divided the total by 6, since there are two extra categories – 'miscellaneous' and 'non-academic'.

Coming back to the differences between the two corpora, one may notice that the most recent texts in COCA are from March 2011, while the most recent texts in the BNC date from the early 1990s, that is, they are older than 15 years. This has major implications for how contemporary English is reflected in the two corpora. In terms of genre balance, the BNC is

10% spoken and 90% written, while in COCA, the genres are evenly divided into 20% for each. Moreover, while in the BNC the spoken section consists of transcripts of recordings of genuine impromptu conversations, the spoken material in COCA mainly contains transcripts of public speeches recorded from the media, a fact which really makes a difference, as the speakers are more self-aware and careful with their language, and hypercorrection may intervene. This, as we shall see, is sometimes reflected in the huge, seemingly inexplicable discrepancies between some of the spoken frequencies obtained from the two corpora. On the other hand, part of the spoken section in COCA is taken from scripts of movies and television series, which in my opinion is an instance of 100% genuine, natural language, as film mimics real life.

For reading convenience, I listed the search strings used in each chapter, in the corresponding tables found in the appendix. Each construction was first compared in the spoken and the written material of each corpus. Then, it was compared in terms of genres (the spoken section was excluded from this stage of analysis). Finally, searches in COHA and Google Books corpora were carried out for a diachronic analysis.

3. Contractions

As already mentioned, Mair (2006: 189) found an increasing use of verb contractions, including negations, in the press sections of the LOB, FLOB, Brown and Frown corpora. It is this aspect of colloquialization – the contracted forms in written language – that this chapter will focus on. Therefore, the question which serves as a starting point for the present section would be: is there an increase in the use of contracted forms in English? If so, what is the current status of contracted forms in the two main varieties of English – British and American? In order to answer these questions, I chose to analyze three categories of verb contractions. A first category is that of standard contractions, that is, those contractions which are accepted as grammatically correct by the current norms of English, e.g. *he'll, I'm, it's, they're, I've, he'd.* A second category is that of negative contractions, also accepted as correct according to the conventions of English, e.g. *can't, won't, wasn't, isn't, don't, didn't, shouldn't, couldn't, wouldn't*. Finally, a third category would be that of what is often referred to as informal, non-standard forms, e.g. *ain't, c'mon, dunnit, dunno, gimme, gonna, gotta, oughta*, or *wanna.* Such non-standard contractions are typically associated with the slang register, and are frequently used in informal speech. The preliminary assumption is that there is an overall increase in the use of contracted forms in English, including the non-standard forms, as a result of a change of language style from formal to informal.

3.1. Standard Contractions

Having set the framework for analysis, we may now proceed to the analysis itself.

Table 1 presents the normalized frequencies of all the contracted forms of *will/shall, have* and *would* in the spoken vs. (the overall) written sections of COCA.

Table 1. Normalized frequencies of verb contractions in the spoken and written material of COCA

	'll	*I'm*	*it's*	*'re*	*'d*	*'ve*
Spoken	1,281.66	1,939.74	3,870.62	3,692.79	602.52	1,743.38
Written	557.50	652.26	1,197.65	873.02	809.14	542.64

The contraction of *be* in the third person singular with nouns and personal pronouns was disregarded, as this is a more difficult to investigate. A search string like '[p*] 's', for instance, will return more than you have bargained for, that is, not only all the lemmas of all types of pronouns plus the contracted form of *be* in the third person singular, but also genitive constructions like *one's work*, *anyone's attempt*, etc. To simplify the analysis, I confined myself to *it's*, a contraction common enough to be representative for this particular verb form. The figures show, as one would expect, that all contractions, with one single exception, are by far more frequent in speech than in writing. The exception is the contracted form of *would* and *had*, *'d*, which, curiously, has a higher frequency in the written than in the spoken section of the corpus. A look among genres (Table 2) shows that in all cases, fiction is at the top of the chart, while academic prose is, as expected, at the bottom of the frequency chart.

Table 2. Normalized frequencies of verb contractions by genre in COCA

Genre	*'ll*	*I'm*	*it's*	*'re*	*'d*	*'ve*
Fiction	1,223.37	1,505.98	1,546.36	1,435.26	2,344.72	915.98
Magazine	579.81	515.79	1,455.18	1,036.84	497.91	609.19
Newspaper	381.32	523.27	1,615.37	921.47	343.14	576.00
Academic	45.52	64.03	173.72	98.54	50.81	69.39

The unusually higher frequency of *'d* in the overall written material than in the spoken section seems to be due to fiction. As Table 2 shows, the frequencies of *'ll, I'm, it's,* and *'re* are more or less similar, but that of *'d* is much higher than the average, almost double the frequency of *'ll* in fiction. At this point, comparing British to American English is of no interest, as there is nothing peculiar about the use of standard contractions in one variety or the other. Diachronically (Figure 1 – Figure 6), all contractions, with the exception of *'ll*, which, curiously, seems to have declined after a frequency peak reached in the 1910s (Figure 1), have an ascending time curve.

Figure 1. Diachronic view of contraction *'ll* in COHA

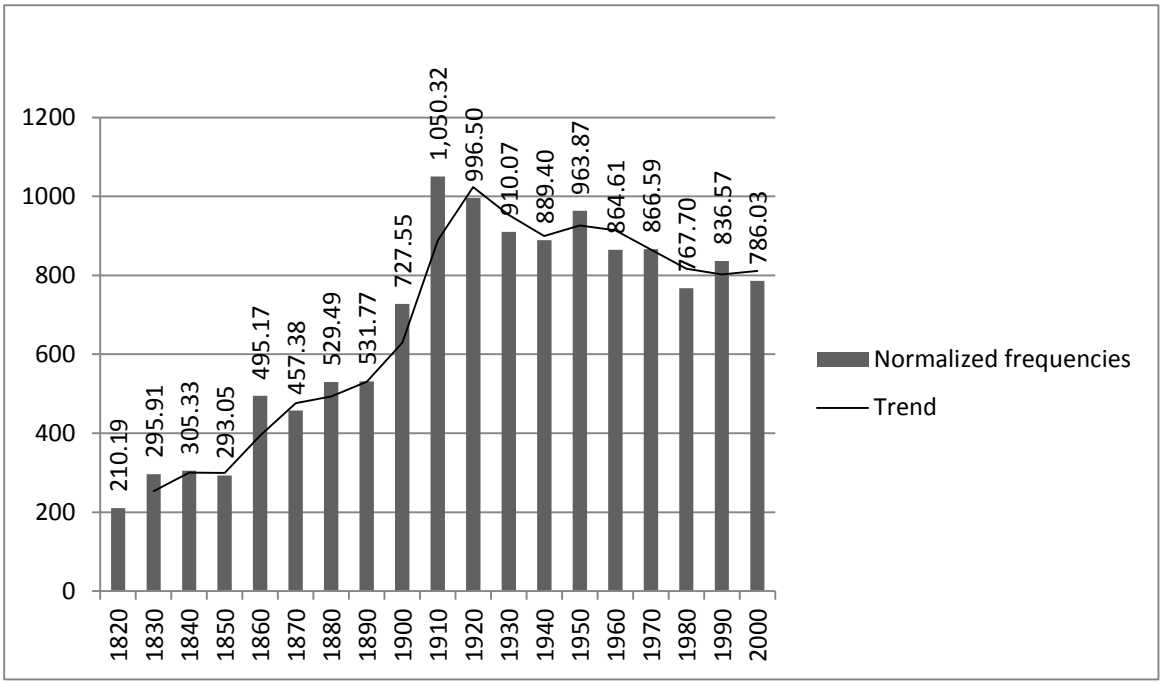


Figure 2. Diachronic view of contraction *I'm* in COHA

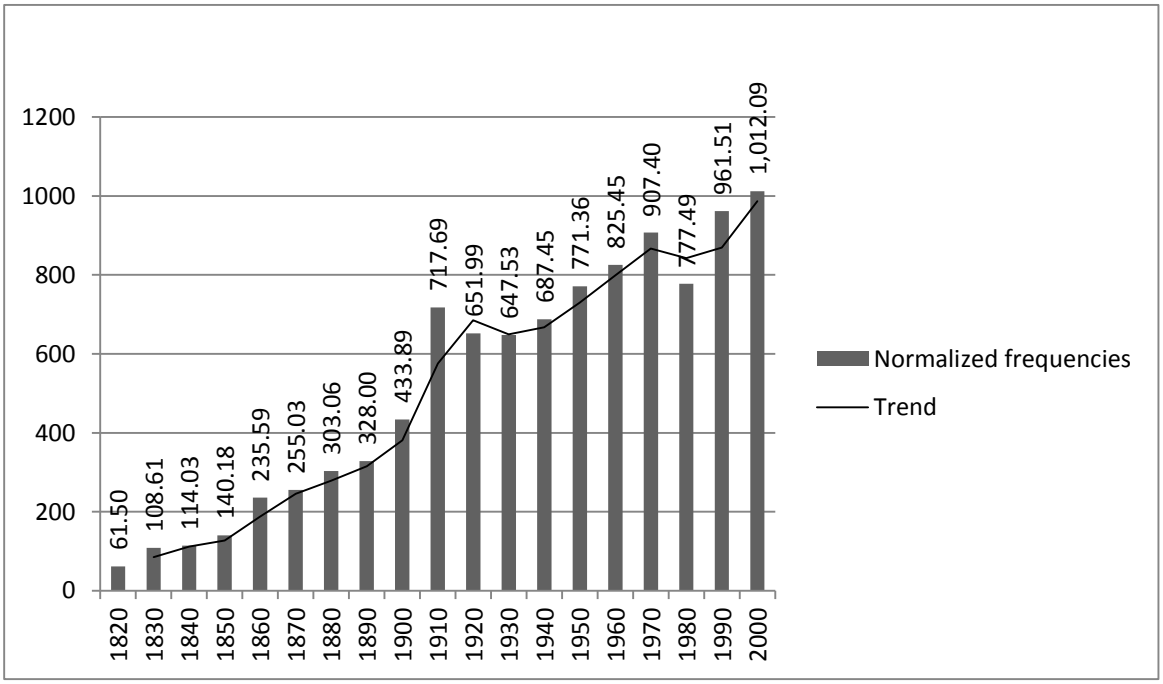


Figure 3. Diachronic view of contraction *it's* in COHA

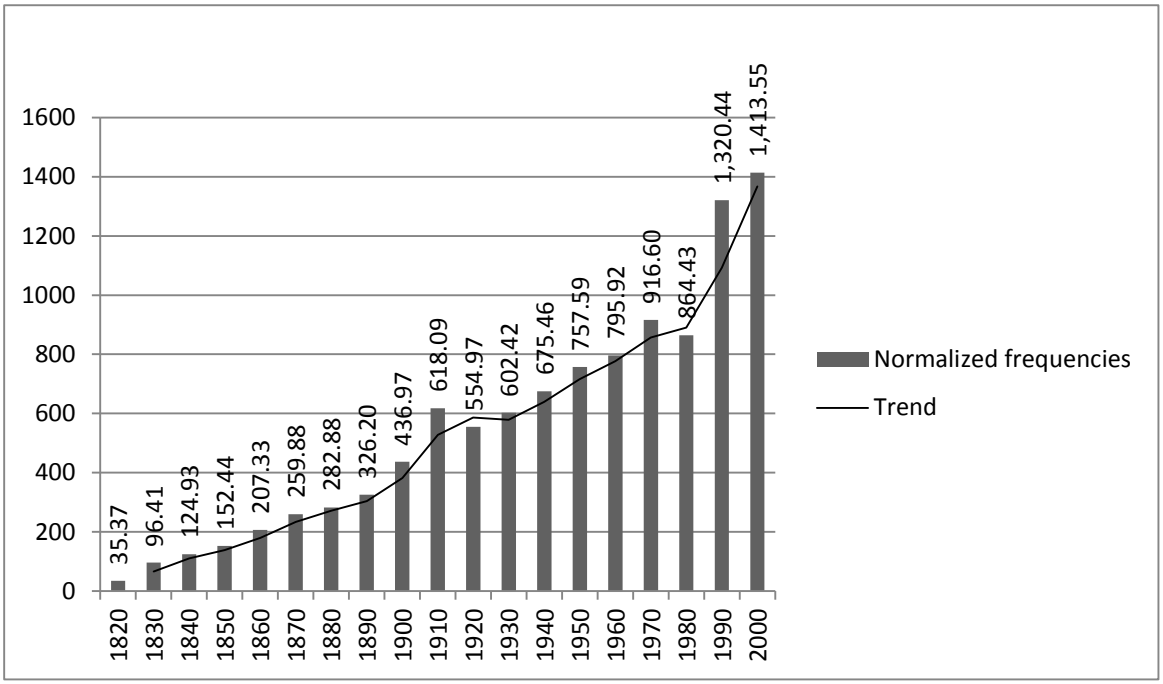


Figure 4. Diachronic view of contraction *'re* in COHA

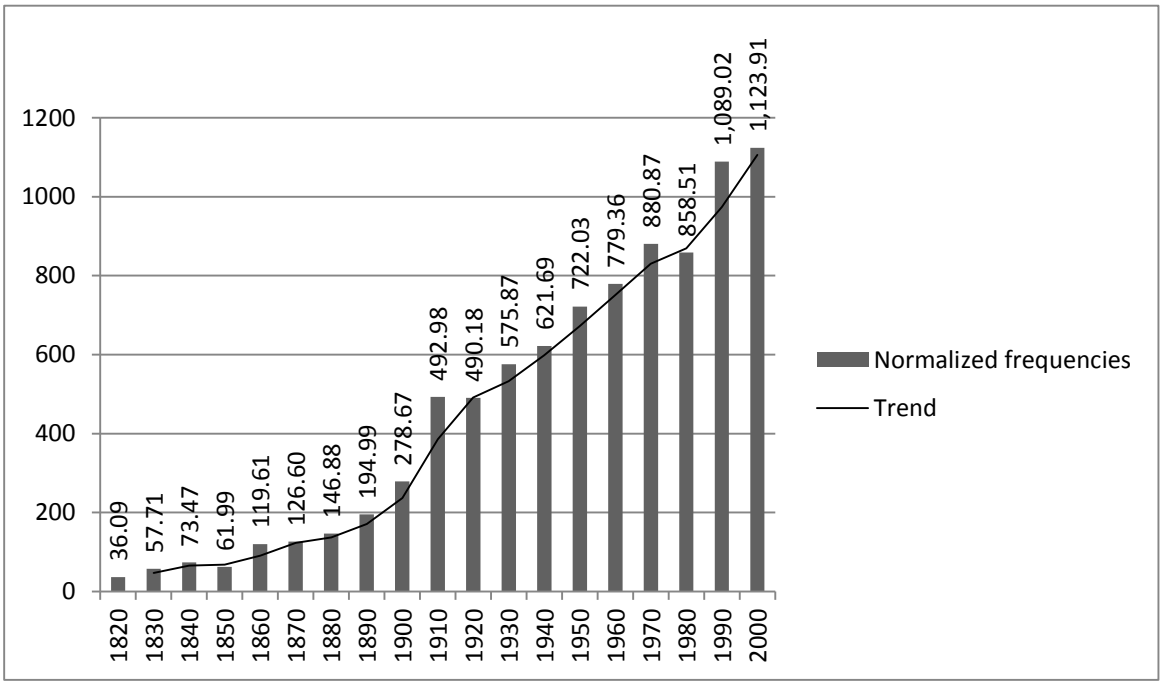


Figure 5. Diachronic view of contraction *'d* in COHA

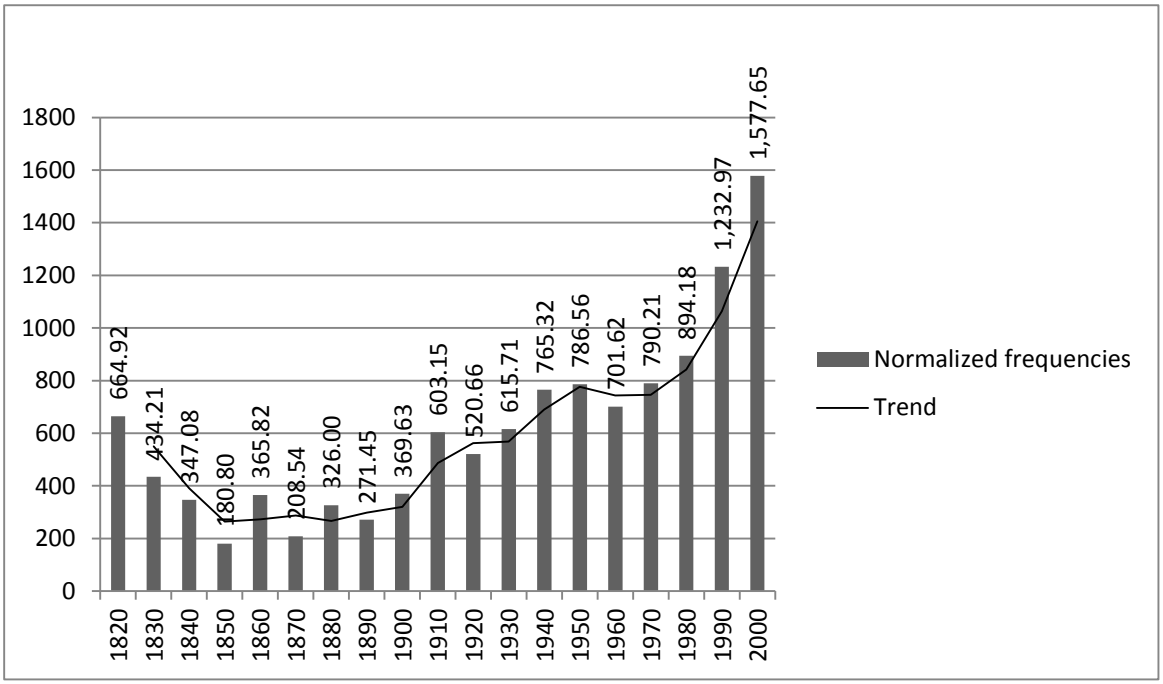
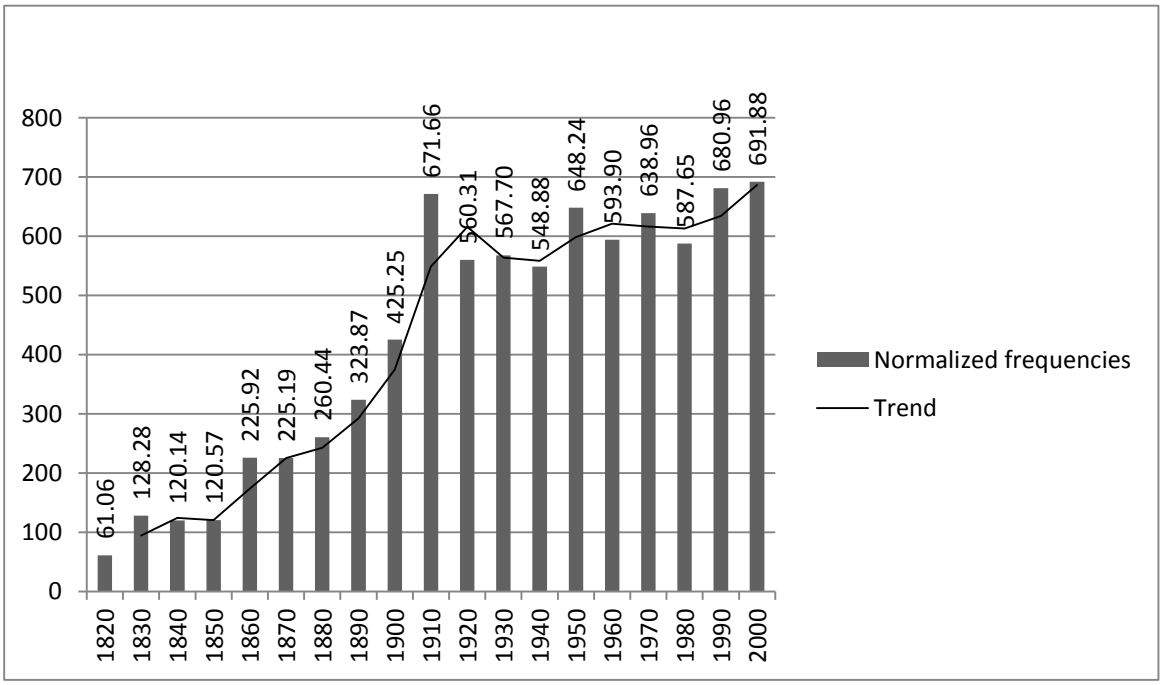


Figure 6. Diachronic view of contraction *'ve* in COHA



A somewhat peculiar development is also seen for the contraction of *would / had* (Figure 5), which, after a decline since the 1820s to the 1850s, has seen a consistent increase up to the present.

3.2.Negative Contractions

As Figure 7 and Figure 8 show, the full form of the negation *cannot* is more frequent in the written sections of the corpus, added up, while the contracted form *can't* is, as expected, more frequent in the spoken section.

Figure 7. Frequencies per million words of *cannot* in the spoken vs. written sections of COCA

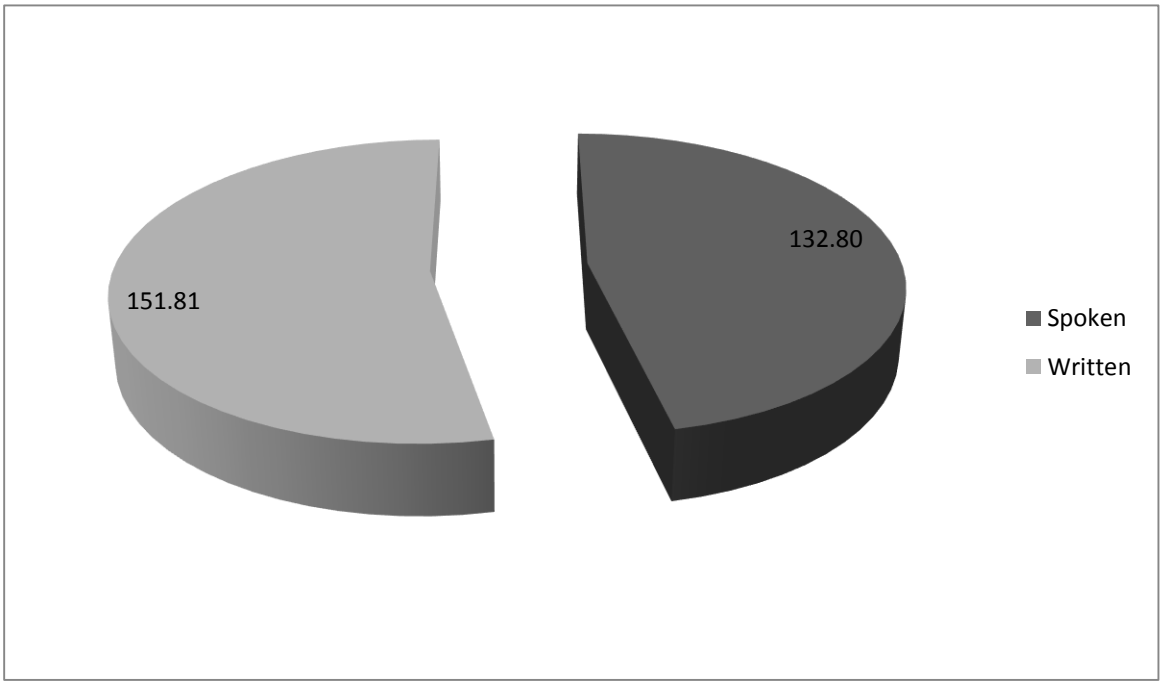


Figure 8. Frequencies per million words of *can't* in the spoken vs. written sections of COCA

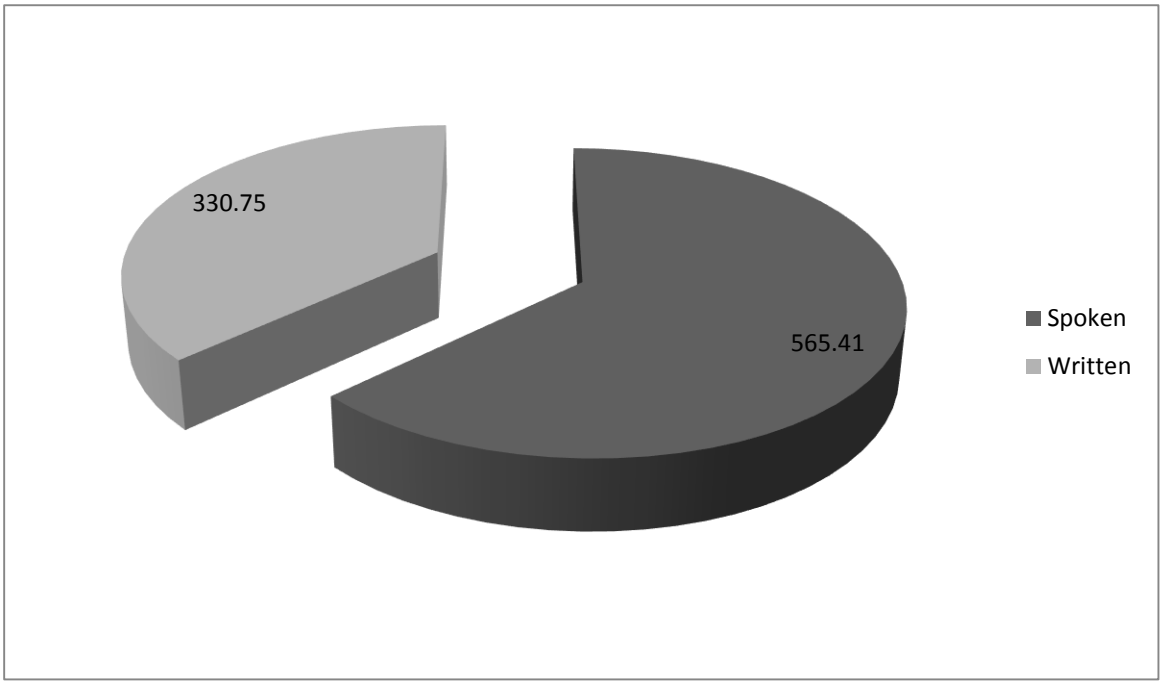
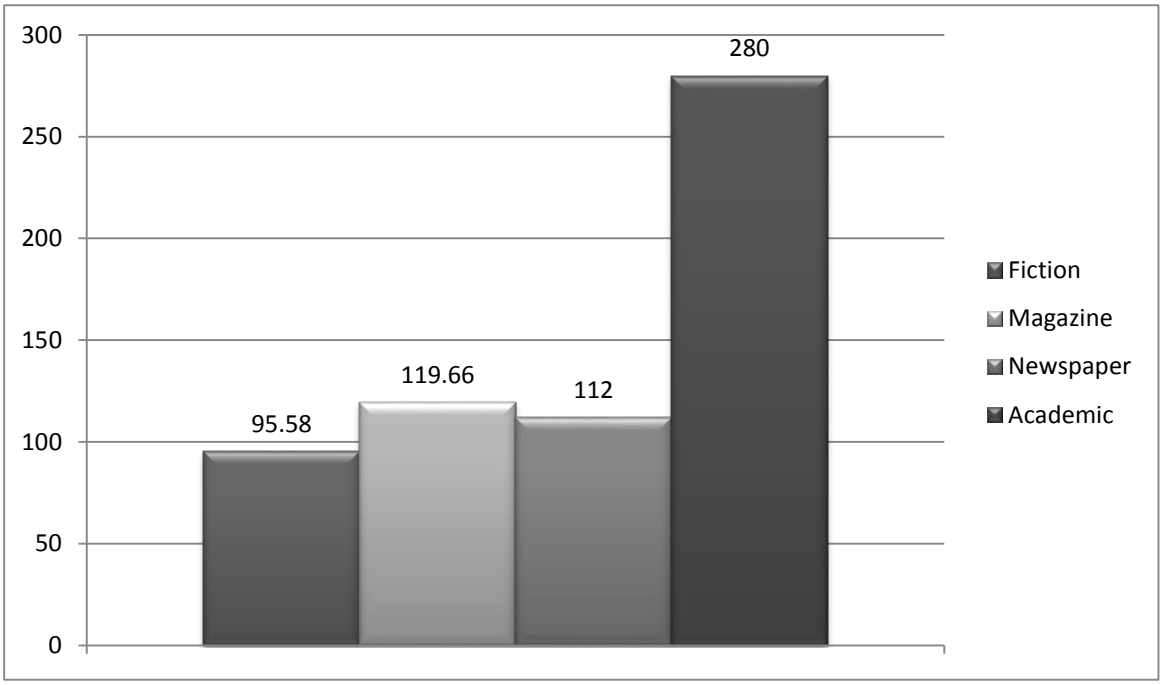


Figure 9. Frequencies per million words of *cannot* in the written material of COCA, divided by genre



A look into genres (Figure 9) shows that academic prose tops the frequency chart of *cannot*, with a normalized frequency of 280. Conversely, the genre chart of *can't* is topped by fiction, with a frequency per million words of 592.95, while academic prose is at the bottom, with a normalized value of only 54.95, as shown by Figure 10.

Figure 10. Normalized frequencies of *can't* in the written material of COCA, divided by genre

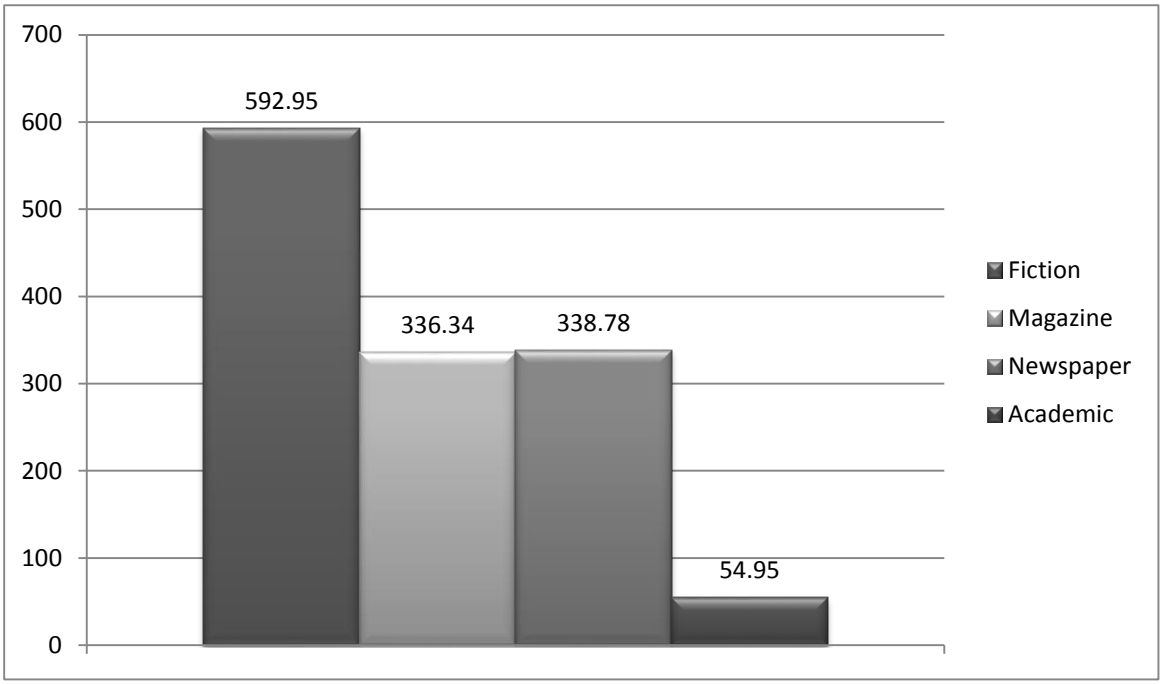
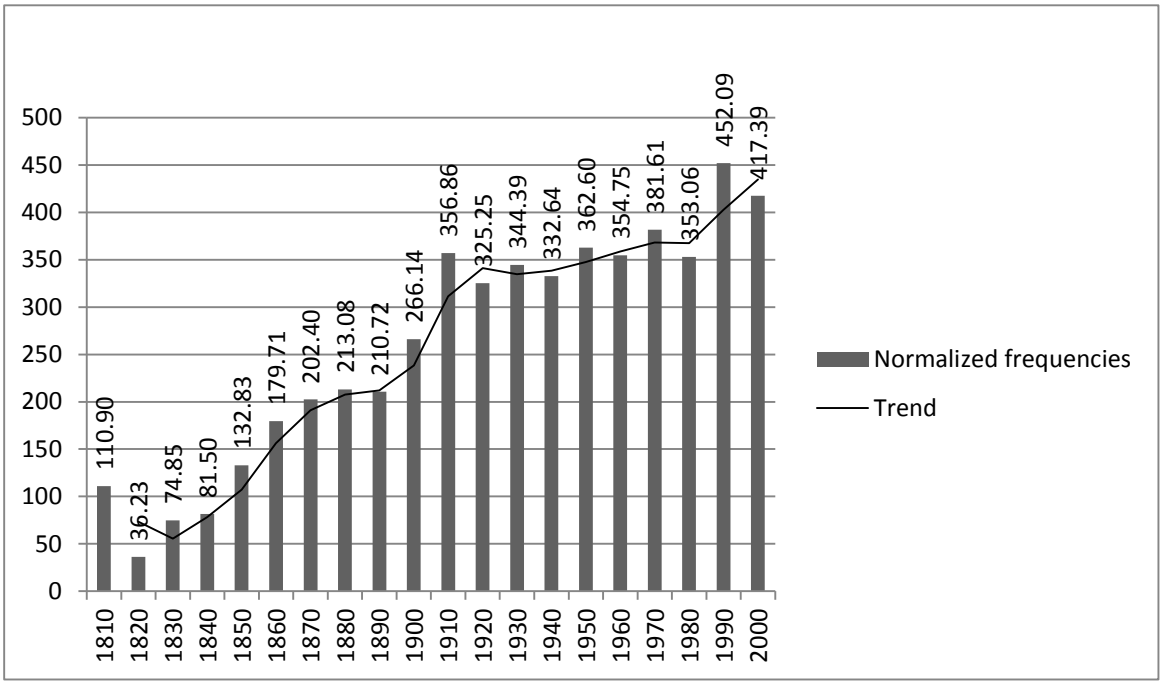


Figure 11. Diachronic view of *can't* in COHA



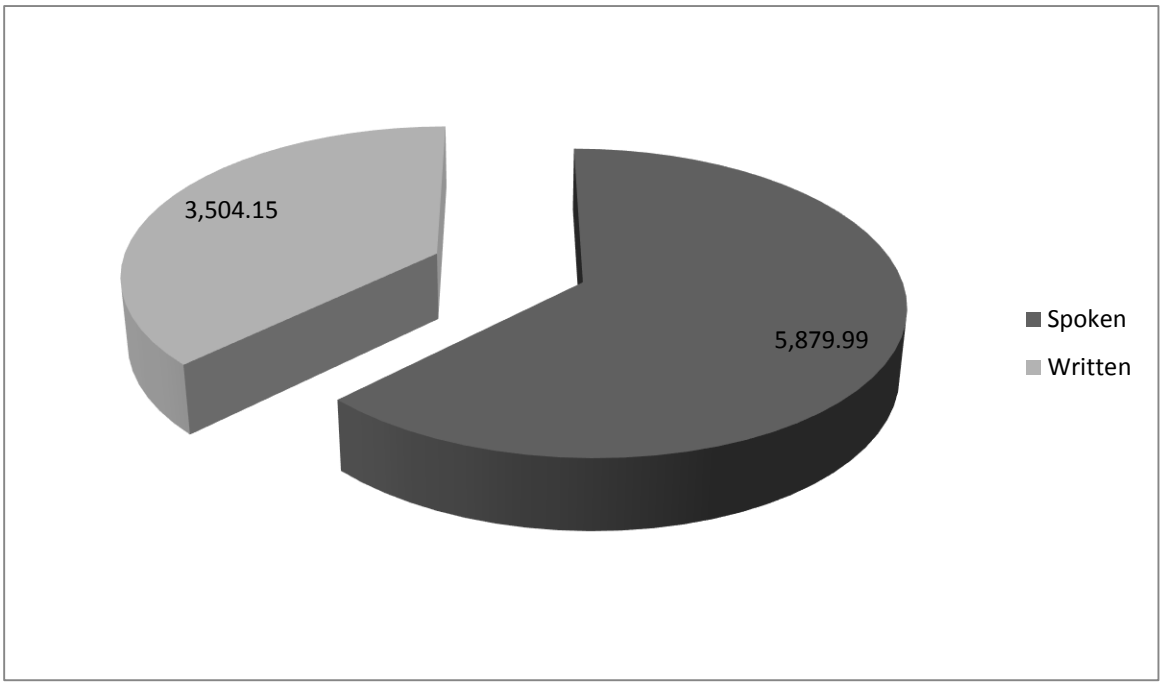
A search for *can't* in COHA shows a clear rising trend of the contracted form since the 1810s (Figure 11). From a manual verification of the instances of *can't* in the academic genre, it is clear that a majority of them are part of a sentence or fragment placed between quotation marks; therefore, the contracted form is not actually part of the academic language proper.

There are, however, a few genuine instances of *can't* in the academic register, as in the example below, found in an academic text in COCA:

(4) However, a religious regime *can't* be rational if it is motivated by religious rewards that can only be claimed in the afterlife. (COCA)

After this concrete example of one of the commonest negative contractions in English, I will now streamline the analysis, searching for all the instances of the negative contraction *n't*. And COCA reveals an exactly identical trend (see Figure 12), with a higher frequency of occurrences in speech than in writing.

Figure 12. Frequencies per million words of *n't* in the spoken vs. written sections of COCA



In terms of genres, the overall use of negative contractions is also similar to that of *can't* (Figure 13), with fiction and academic prose at the top and bottom of the chart, respectively. Figure 14 shows a steady increase in the use of negative contractions since the 1820s to the present.

Figure 13. Normalized frequencies of *n't* in the written material of COCA, divided by genre

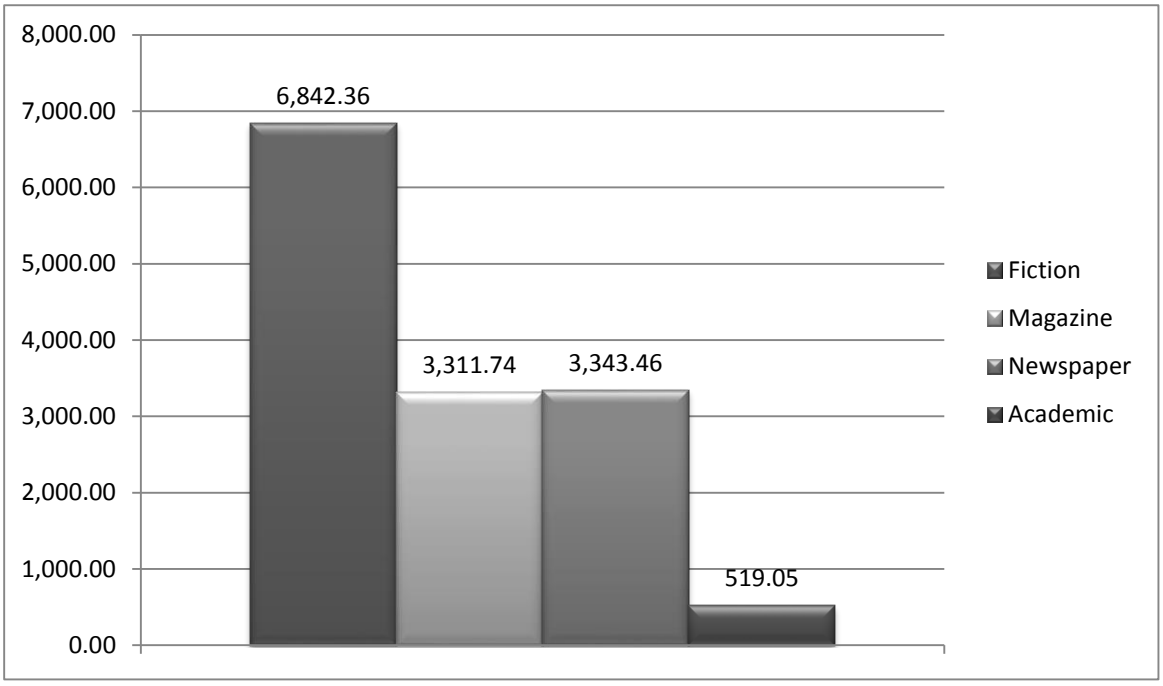
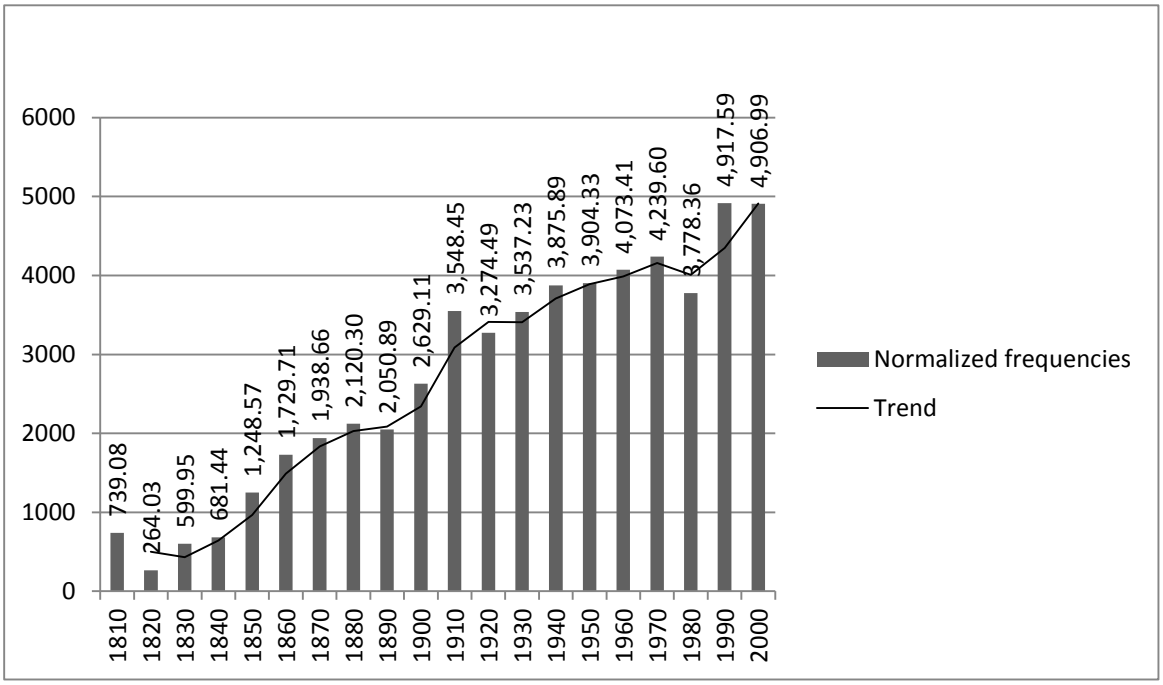


Figure 14. Diachronic view of *n't* in COHA



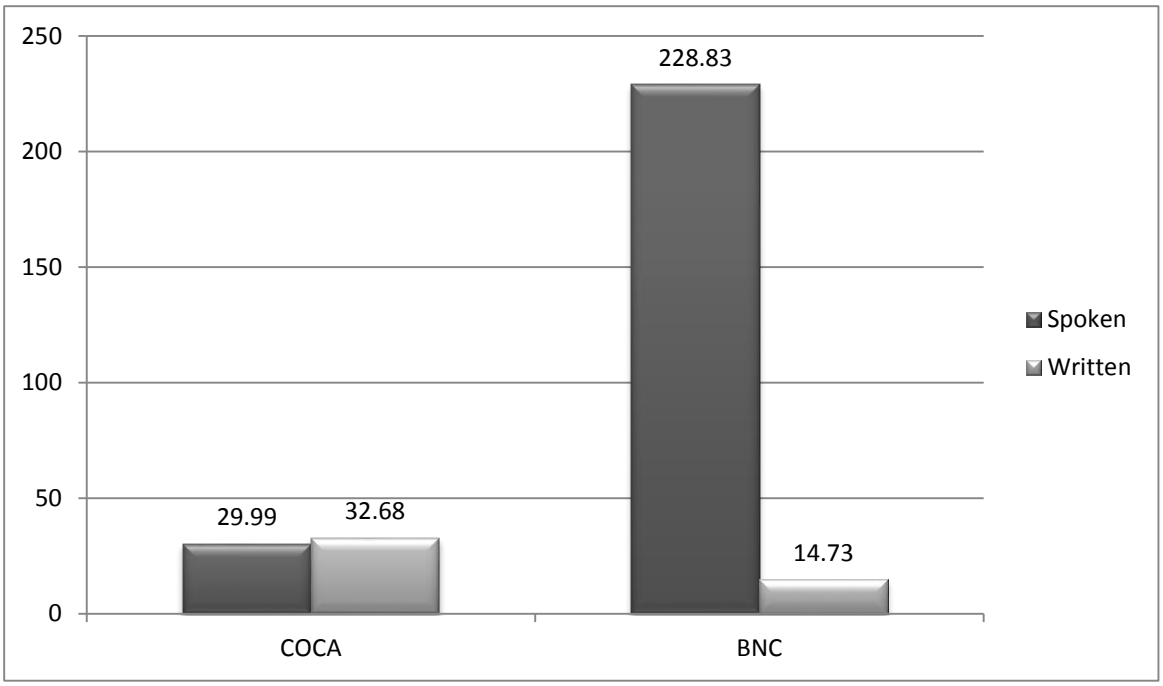
3.3. Non-Standard Contractions

According to *The Concise New Partridge Dictionary of Slang and Unconventional English* (2008: 5), *ain't* or *aint* originates in the UK, in 1710, and is a slang contraction of *am not, are not, is not, has not, have not*.

(5) Give it up, Lieutenant. They *ain't* caught the guy in the last thirty years, they *ain't* gonna catch him in the next ten days. (COCA)

With a normalized value of 32.68, *ain't* is more frequent in writing than in speech in COCA, the frequency per million words in the spoken section being 29.99, as seen in Figure 15 below.

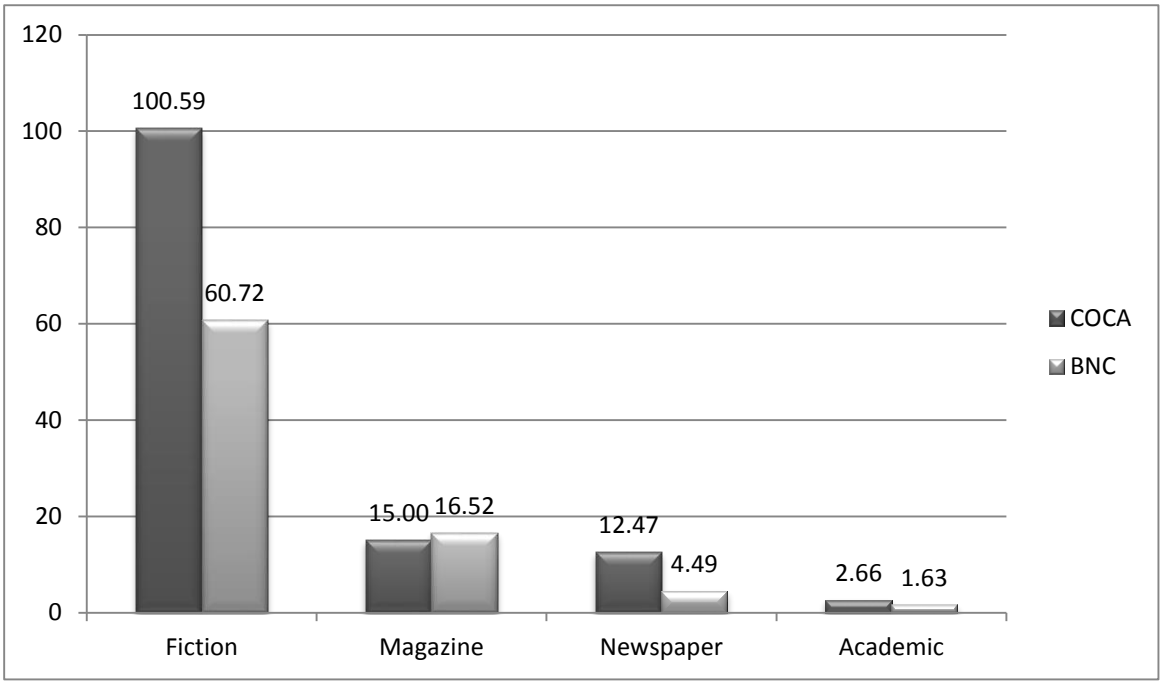
Figure 15. Frequencies per million words of *ain't* in COCA and BNC in spoken vs. written material



On the other hand, the difference between the frequencies in the spoken and written sections of BNC is huge, with a much higher frequency in the spoken section. In writing, the contraction is twice as frequent in COCA as in the BNC, as Figure 15 shows.

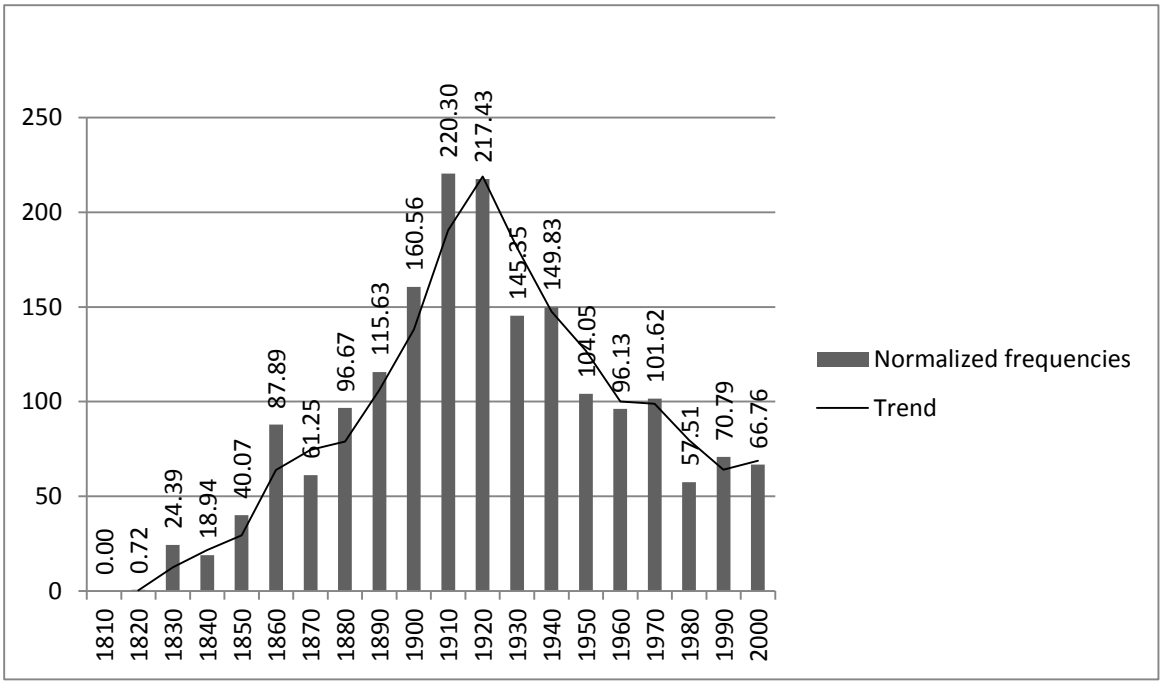
A look at frequencies by genres (see Figure 16 below) shows similar proportions in the two corpora, with fiction at the top, and academic prose at the bottom of the chart.

Figure 16. Frequencies per million words of *ain't* in the written material of COCA and BNC, divided by genre



The samples found in the academic category of both corpora are parts of quotations, therefore, like negative contractions, *ain't* does not seem to be actually part of the academic language proper. A diachronic look at *ain't* in COHA shows an increasing use of this contraction since the 1820s up to the 1910s, but a firm decline after this peak, up to the present (see Figure 17).

Figure 17. Diachronic view of *ain't* in COHA



Dunno, *dunna* or *dunnaw* are slang contractions of the standard negative contraction of the verb *know* (*don't know*); it first occurred in the UK, in 1812 (cf. *The Concise New Partridge Dictionary of Slang and Unconventional English* 2008: 226).

(6) "Tell me, Constable," she said, sipping her drink with maidenly primness, "the small gentleman lunching with your sergeant. He is not, I think, quite English?" "*Dunno*." The constable's speech was impeded by corned beef sandwich. When he swallowed he said, "Leastways he bain't from these parts." (BNC: A0D)

The frequency per million words of *dunno* in COCA is higher in the spoken than in the written section, while in the BNC, the frequency in the spoken section is much higher than that in writing, as seen in Figure 18.

Figure 18. Frequencies per million words of *dunno* in COCA and BNC in spoken vs. written material

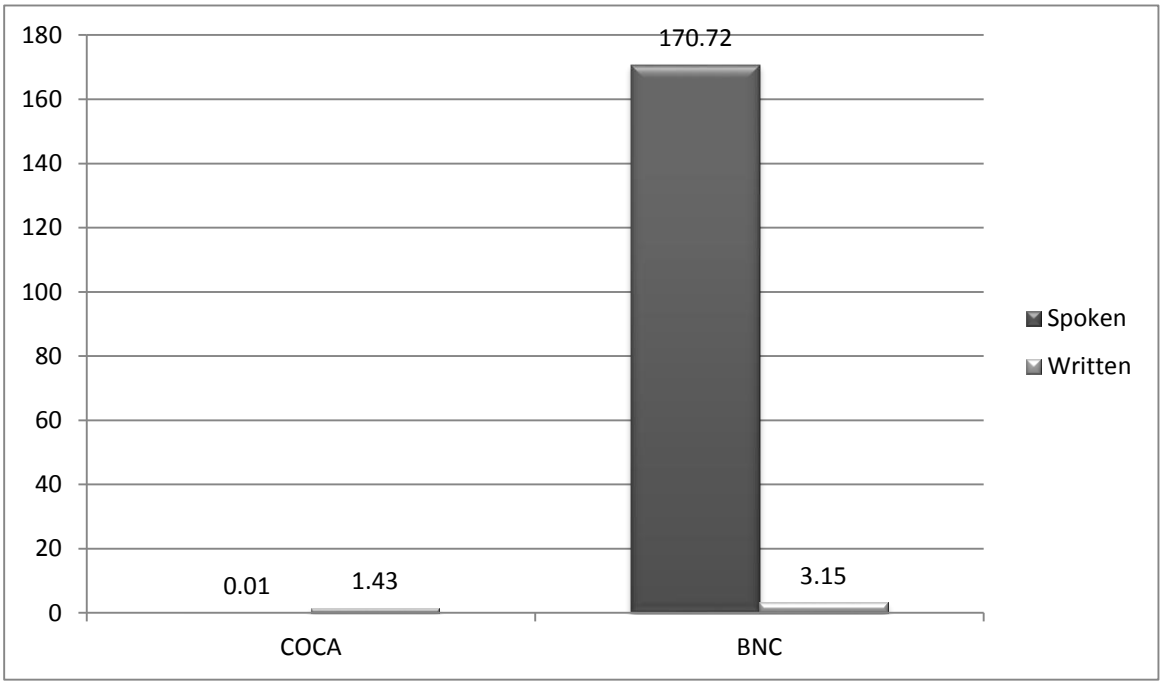
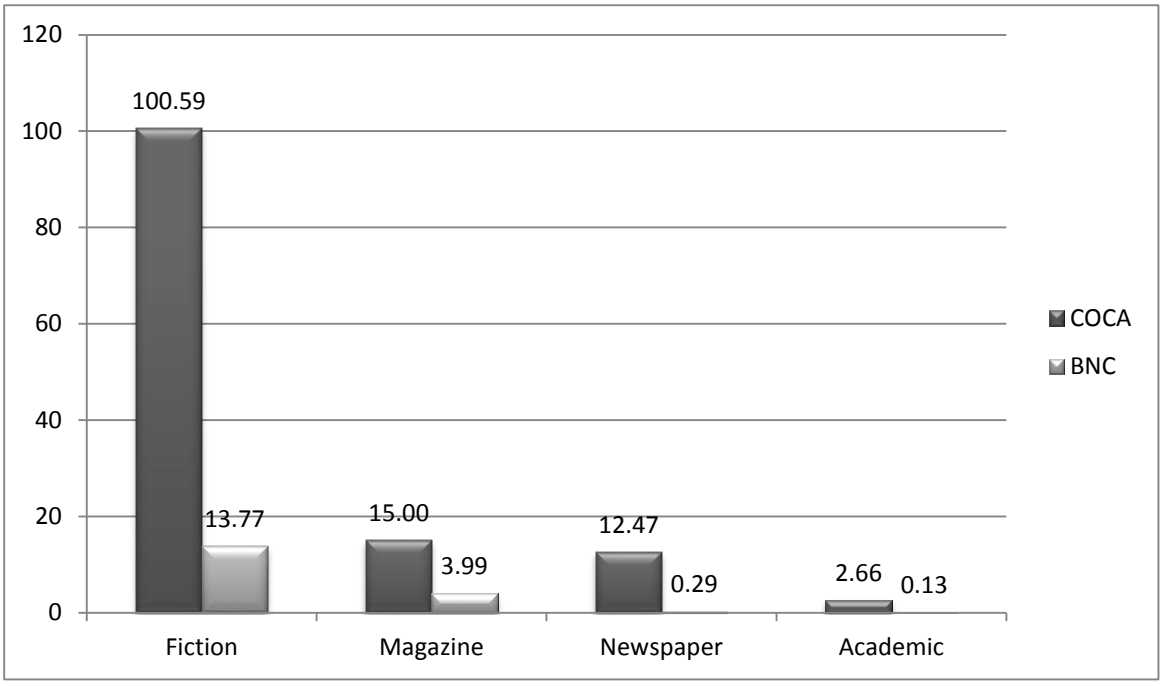
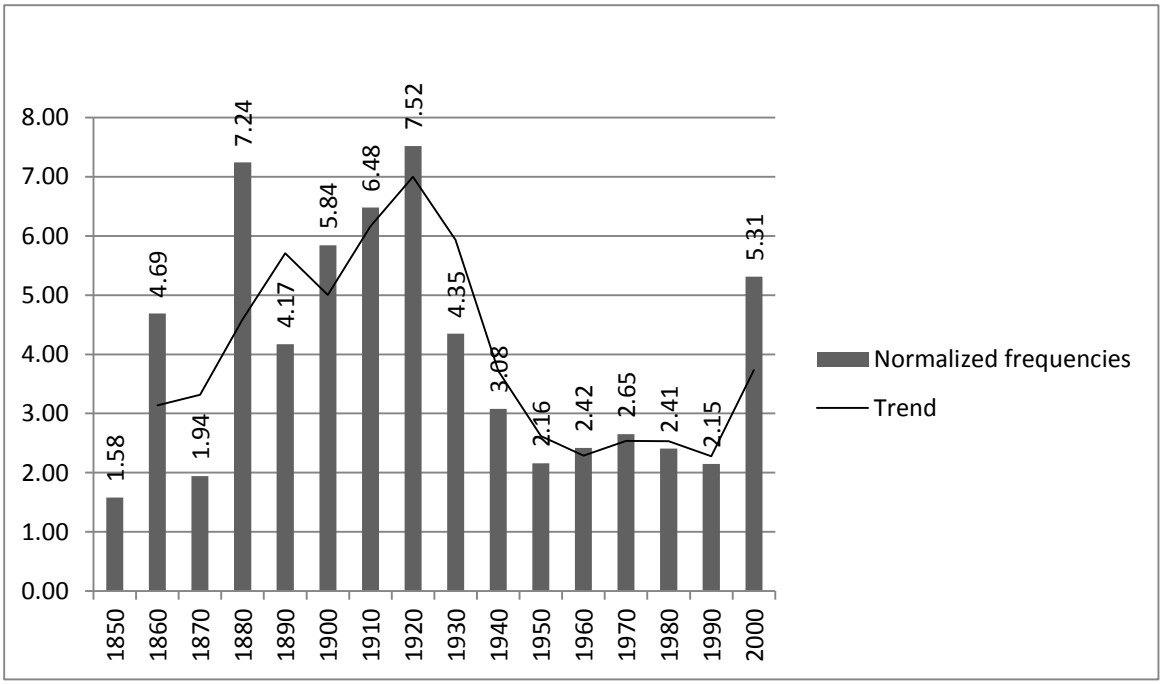


Figure 19. Frequencies per million words of *dunno* in the written material of COCA and BNC, divided by genre



In terms of genres, fiction is, as usual, at the top of the chart, while academic writing, at the bottom, as Figure 19 shows. The diachronic evolution of *dunno* in COHA is indecisive, with ups and downs at various intervals, the last trend being a seemingly ascending one, as seen in Figure 20.

Figure 20. Diachronic view of *dunno* in COHA

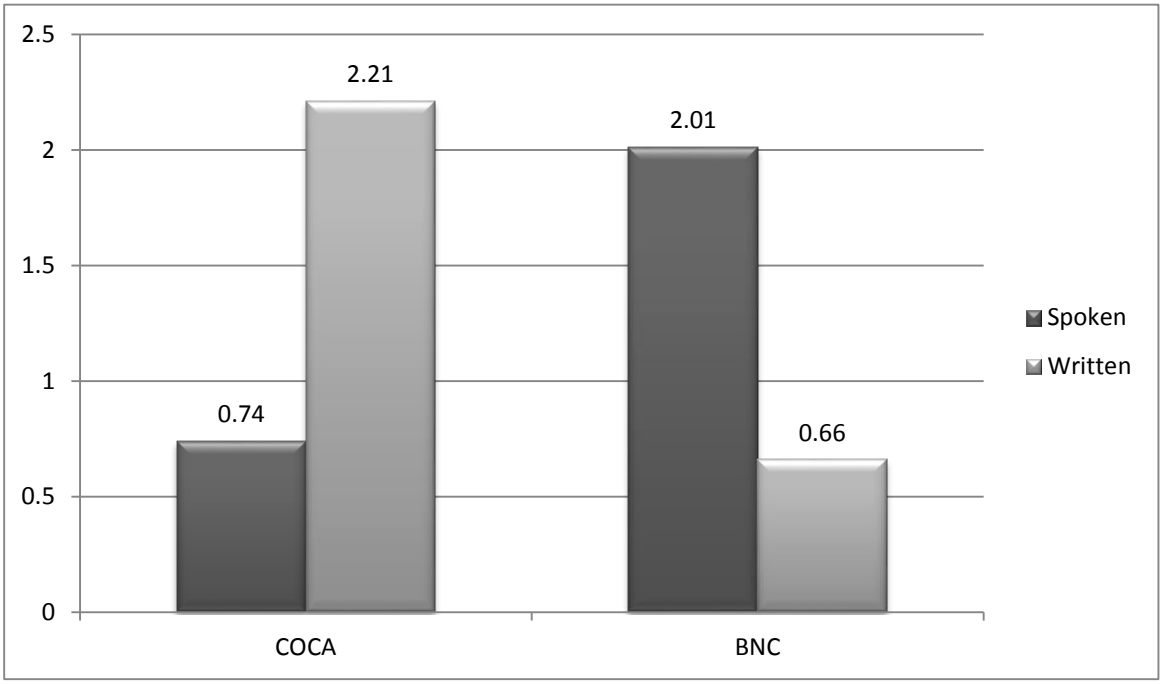


According to the same *Concise New Partridge Dictionary of Slang and Unconventional English* (2008: 289), *gimme* is 'a lazy phonetic abbreviation' of *give me*, originated in the US, in 1883.

(7) "So anyway," he says airily, opening it. "Let's see what we've got here. *Gimme* just a sec." "All right," she says. (COCA)

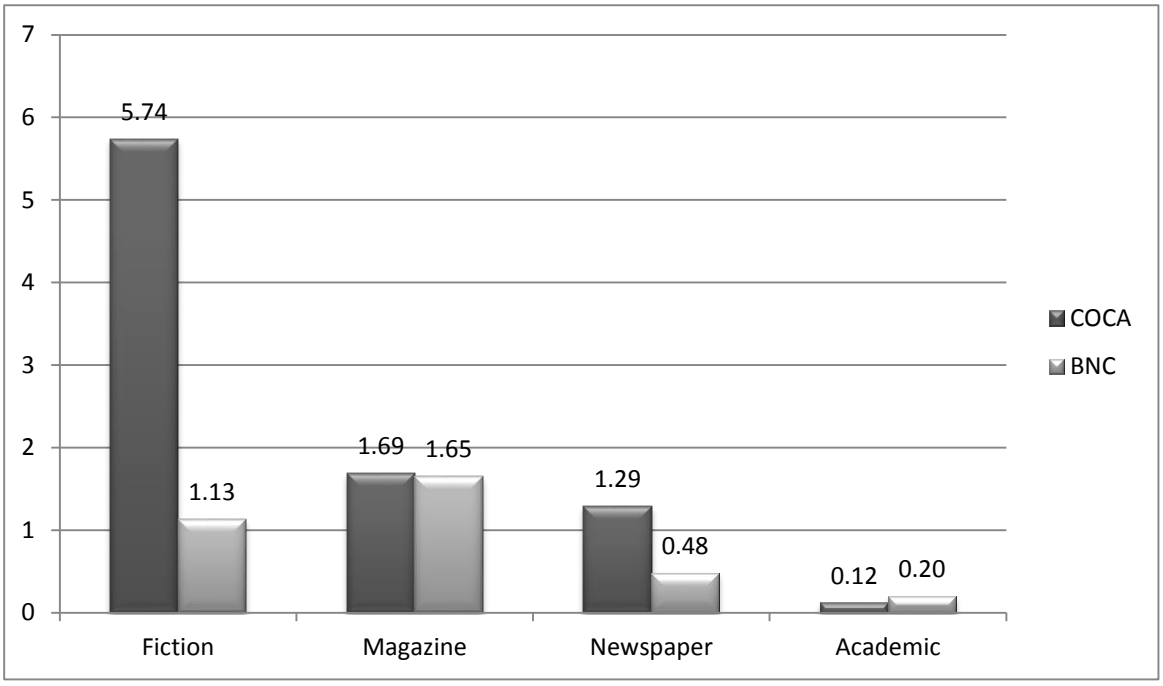
Figure 21 shows that the normalized frequency of *gimme* in the written material of COCA is much higher than that in the spoken section of the same corpus.

Figure 21. Frequencies per million words of *gimme* in COCA and BNC in spoken vs. written material



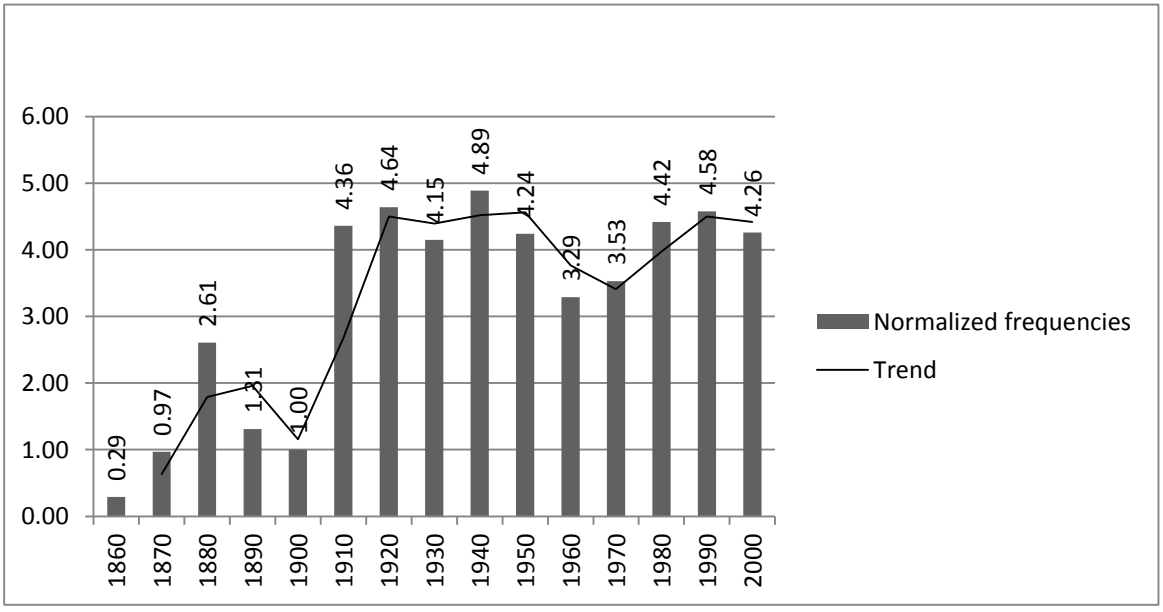
Conversely, the BNC displays an opposite proportion between its spoken and written sections, with a normalized frequency of 2.01 in speech, and only of 0.66 in writing. The reason for such discrepancies between the two corpora and between the spoken and written sections of each of them might be their different composition. In terms of genres, fiction is, again, at the top of the chart in COCA (as seen in Figure 22 below), followed, in descending order, by the magazine, newspaper and academic genres.

Figure 22. Frequencies per million words of *gimme* in the written material of COCA and BNC, divided by genre



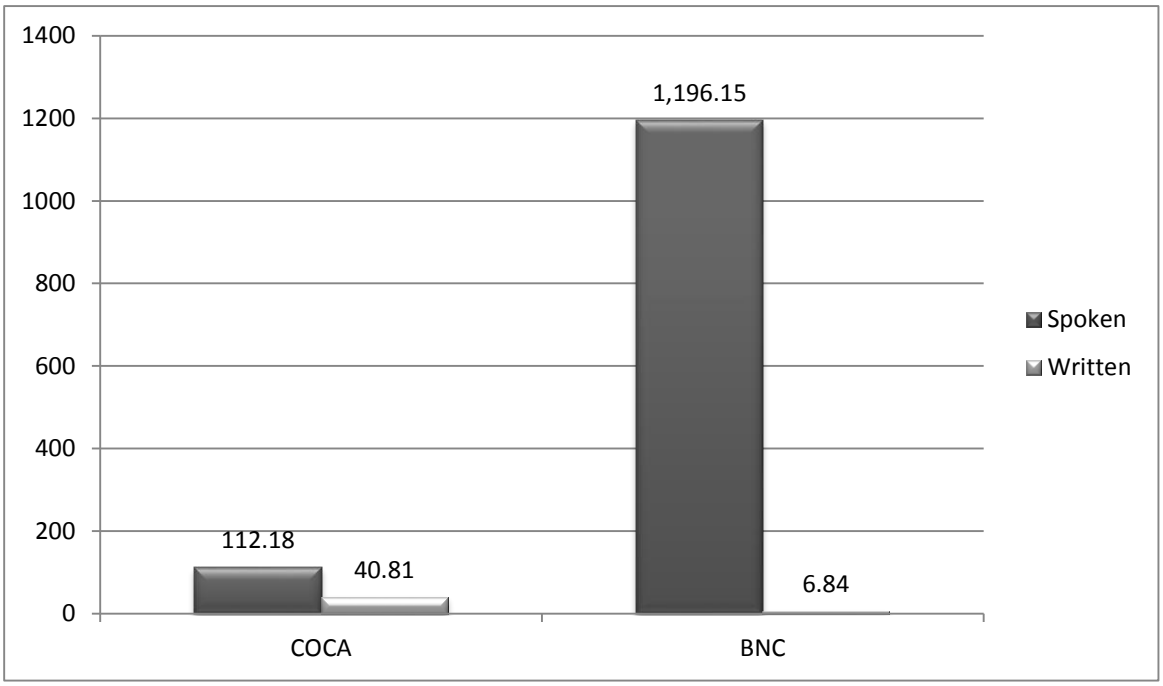
Things are slightly different this time in the case of the BNC, where the highest frequency is reached by the magazine category, followed by fiction, newspaper and academic. Like *dunno*, the historical trend of *gimme* in COHA is uncertain, with ups and downs at times, and with a declining tendency during the latest period included in the corpus, namely between 1990 and 2009 (see Figure 23).

Figure 23. Diachronic view of *gimme* in COHA



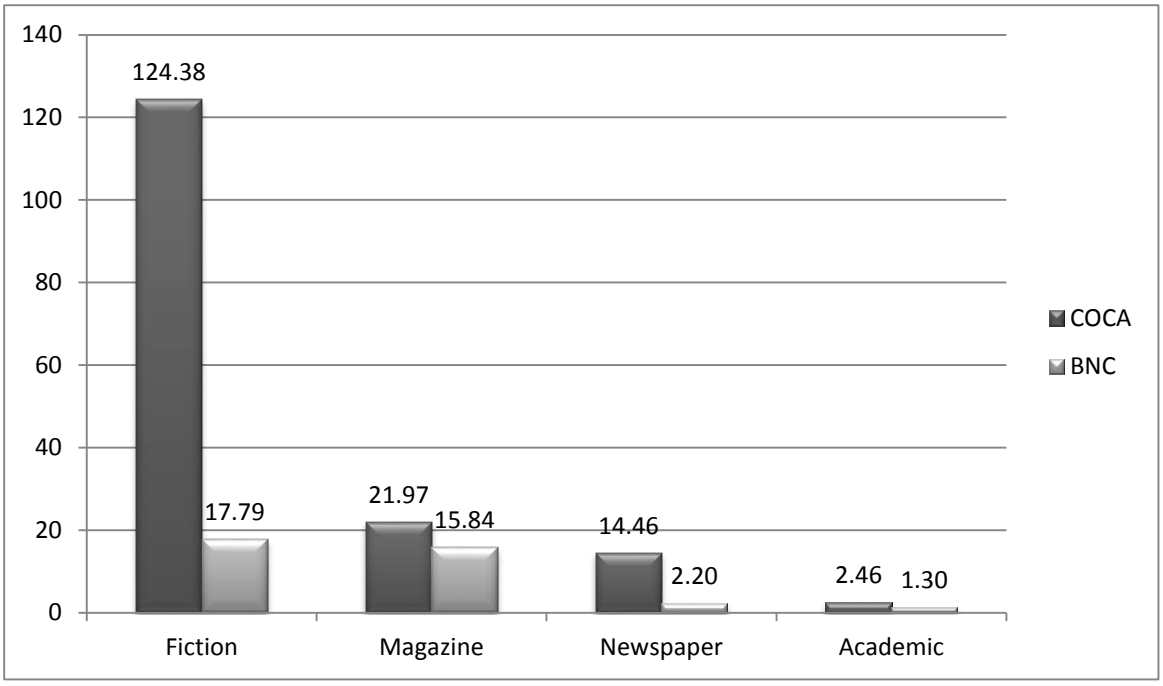
In his discussion about the *be going to* + Infinitive construction as a way of referring to future time in English, Leech (2004: 58) notes that informally, the pronunciation of *going to* is reduced to /ˈgɔnə/; hence the non-standard spelling *gonna*. Mair (1997) has already pointed out, after his corpus-based study, that the *be going to*-future is spreading both in British and American English (cf. Mair 1997: 173), but how about its contracted form *gonna*? In both corpora, the normalized frequencies of this contraction are much higher in the spoken section than in the written material, but the frequency in the American written material is also much higher than that in the British written sections, which confirms *gonna* as a typical American expression (see Figure 24).

Figure 24. Frequencies per million words of *gonna* in COCA and BNC in spoken vs. written material



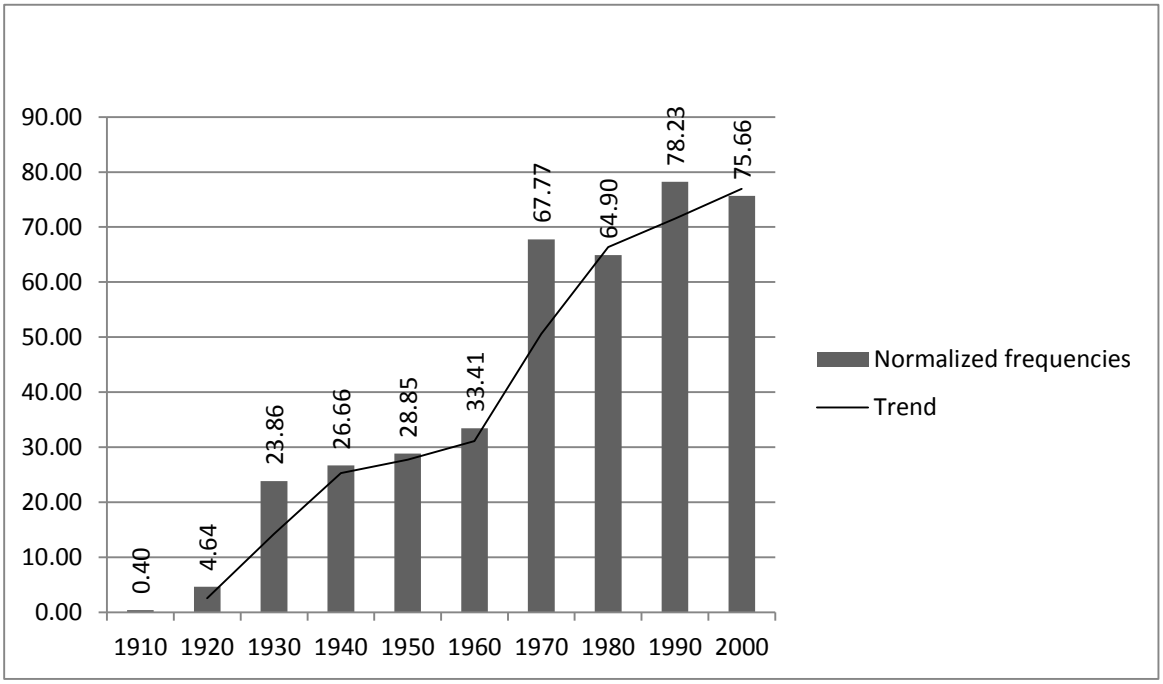
In terms of genres, both corpora have the same order of frequencies, with the highest in fiction, followed, in order, by the magazine, newspaper, and academic genres (Figure 25).

Figure 25. Frequencies per million words of *gonna* in the written material of COCA and BNC, divided by genre



The diachronic analysis presented in Figure 26 clearly shows a marked increase in the use of *gonna* since its first occurrence in COHA, in the 1910s onwards.

Figure 26. Diachronic view of *gonna* in COHA

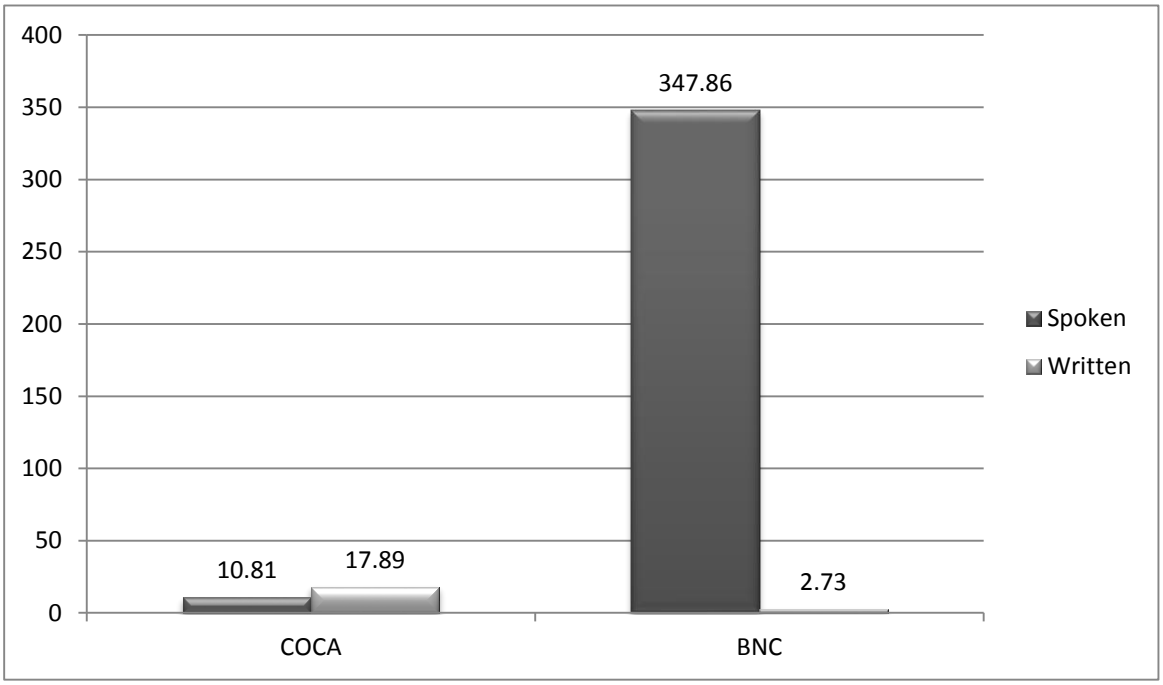


Gotta is the contracted form of *got to* meaning *have to*.

(8) "Sorry," she said. "I *gotta* go. Thanks, mister. I don't want trouble. Good night;" She kissed me on the cheek, and followed the three out to the street. (BNC: GVL)

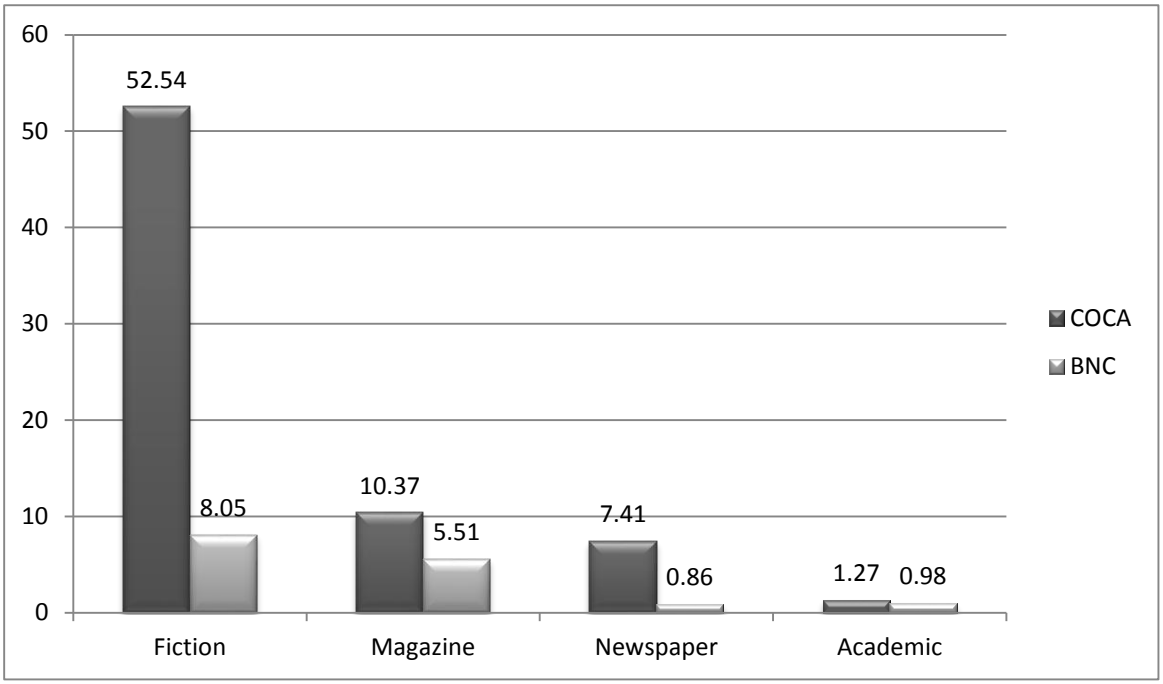
The normalized frequency of *gotta* in the spoken material of COCA is 10.81, while in the BNC it is 347.86 (see Figure 27). This huge discrepancy must be, as already stated, due to the very different composition of the two corpora. The frequency of *gotta* in the written material of COCA is higher than that of the spoken section, whereas in the BNC, the number of occurrences in the spoken section is much larger than that in the written material of the corpus. The instances of *gotta* in the written sections of COCA are much more numerous than those in the written material of the BNC.

Figure 27. Frequencies per million words of *gotta* in COCA and BNC in spoken vs. written material



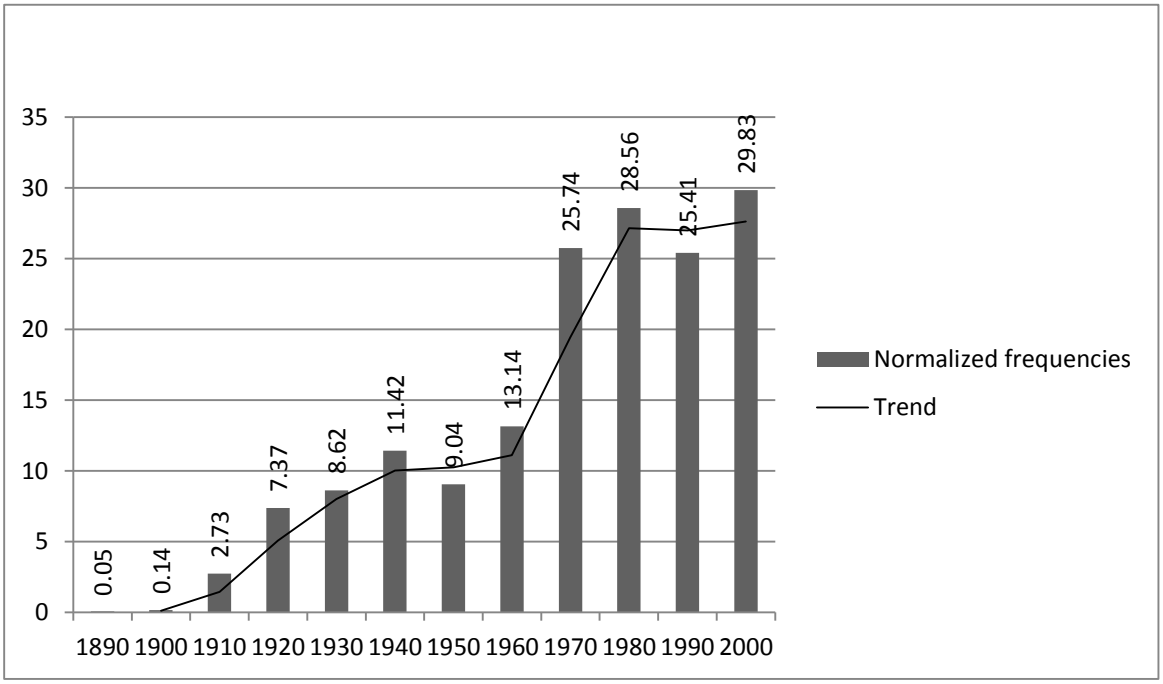
In terms of genres (Figure 28), the frequencies follow the general pattern, with fiction at the top of the chart, followed, in order, by the magazine, newspaper, and academic categories.

Figure 28. Frequencies per million words of *gotta* in the written material of COCA and BNC, divided by genre



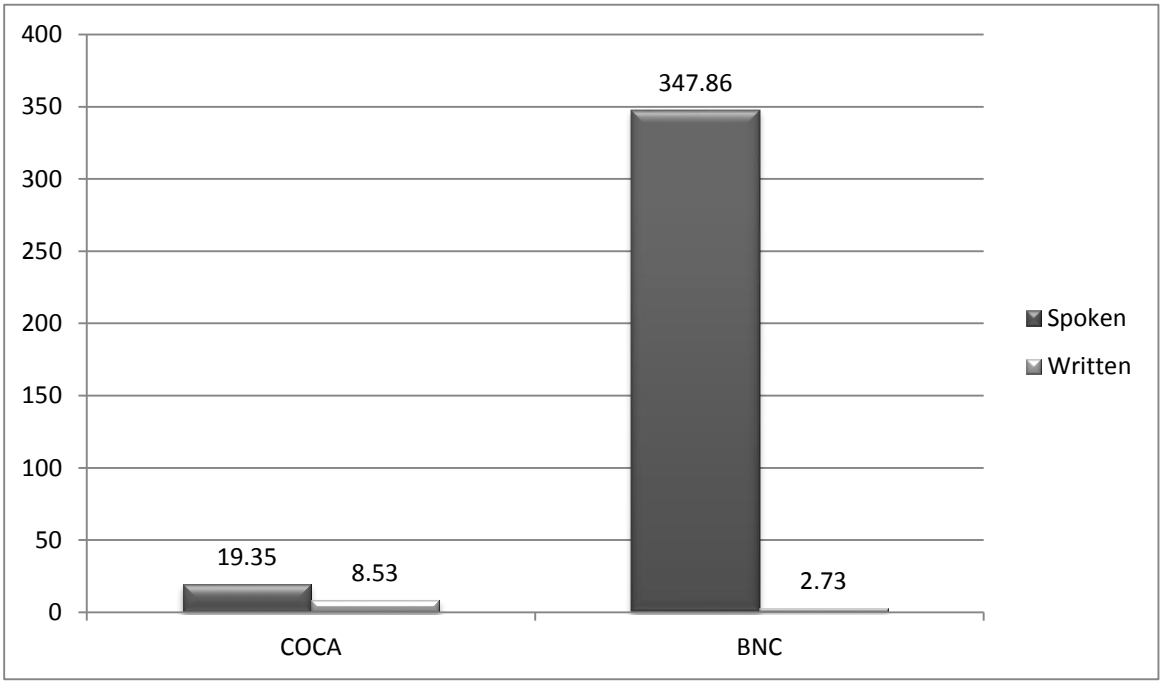
Like *gonna*, and unlike *ain't, dunno* and *gimme*, *gotta*, displays a clear ascending diachronic path in COHA, as seen in Figure 29.

Figure 29. Diachronic view of *gotta* in COHA



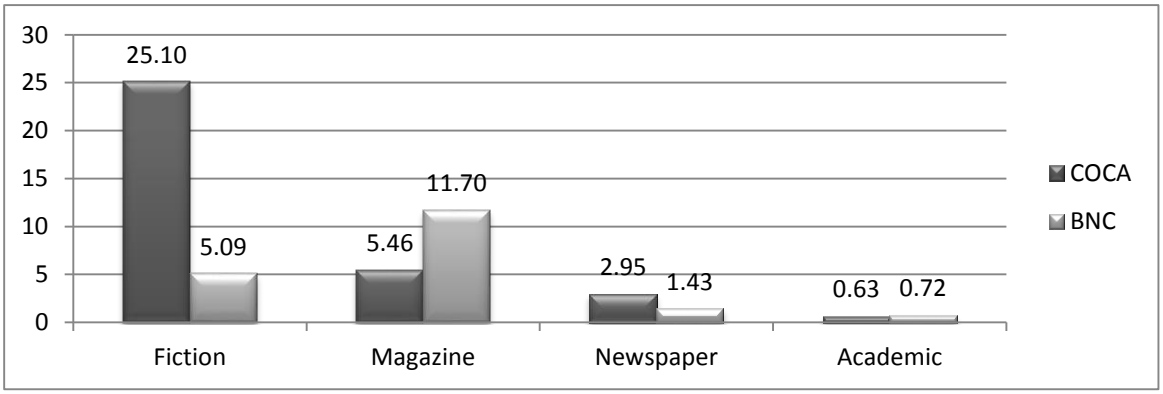
Wanna is the contracted form, in the slang register, of *want to.* In both COCA and the BNC the frequencies in the spoken sections are higher than the frequencies in the written material (see Figure 30).

Figure 30. Frequencies per million words of *wanna* in COCA and BNC in spoken vs. written material



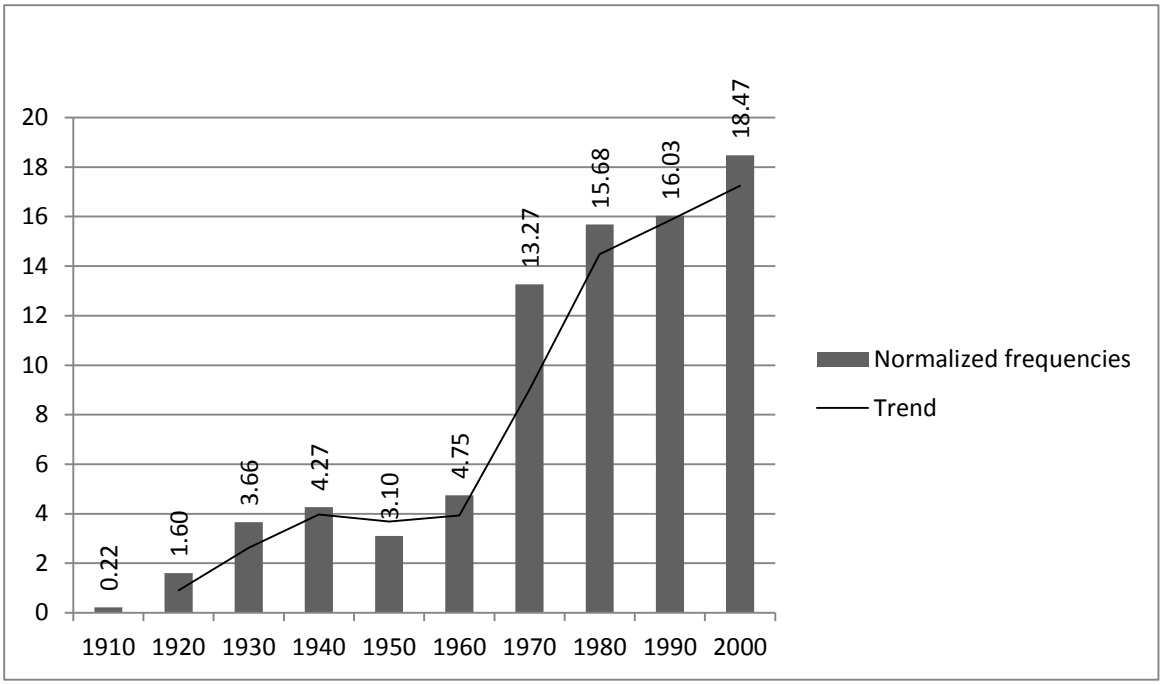
On the other hand, the frequency in the spoken section of BNC is much higher than the equivalent frequency in COCA, while the frequency in the written material of COCA is higher than that in the written material of the BNC. In terms of genres, COCA has the highest frequency of *wanna* in fiction, followed by the magazine, newspaper, and academic genres (Figure 31).

Figure 31. Frequencies per million words of *wanna* in the written material of COCA and BNC, divided by genre



The BNC instead, has the highest frequency in the magazine category, followed by fiction, newspapers, and academic prose. The frequencies in the academic genre are roughly similar in the two corpora, with 0.63 in COCA and 0.72 in the BNC (normalized values). Diachronically, *wanna* records the same sharp increase in COHA as *gonna* and *gotta*, since 1910 to the present (Figure 32).

Figure 32. Diachronic view of *wanna* in COHA

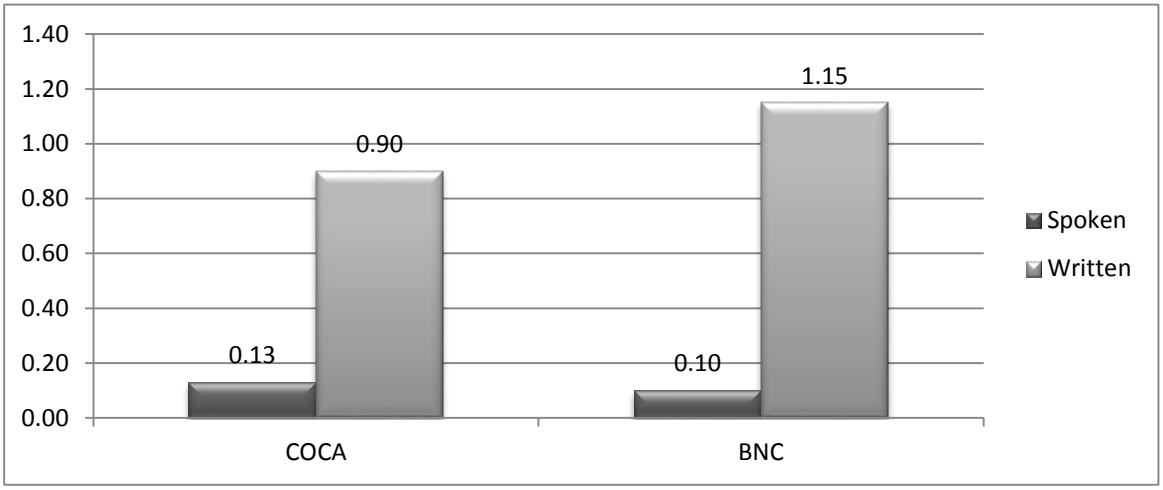


3.4. Other Contractions

Although not a verb contraction but still a contraction, *lotta*, along with its less known variant *lorra*, is described by *The Concise New Partridge Dictionary of Slang and Unconventional English* (2008: 409) as a 'slovening' of the determiner *lot of*, which first occurred in the UK, in 1906. It has an American equivalent, *lotsa*, a slovenly contraction of the more informal *lots of*, first found in the US in 1927 (cf. *The Concise New Partridge Dictionary of Slang and Unconventional English* 2008: 409). In order to find out more details about the use of *lotta* in the two main varieties of English, I first performed searches in COCA and the BNC, treating, as usual, the written material as a whole. To eliminate the personal name *Lotta* from the search results, I used *a lotta* as the search string instead of the single word, although I was fully aware that thus I would miss all constructions including an adjective or adverb

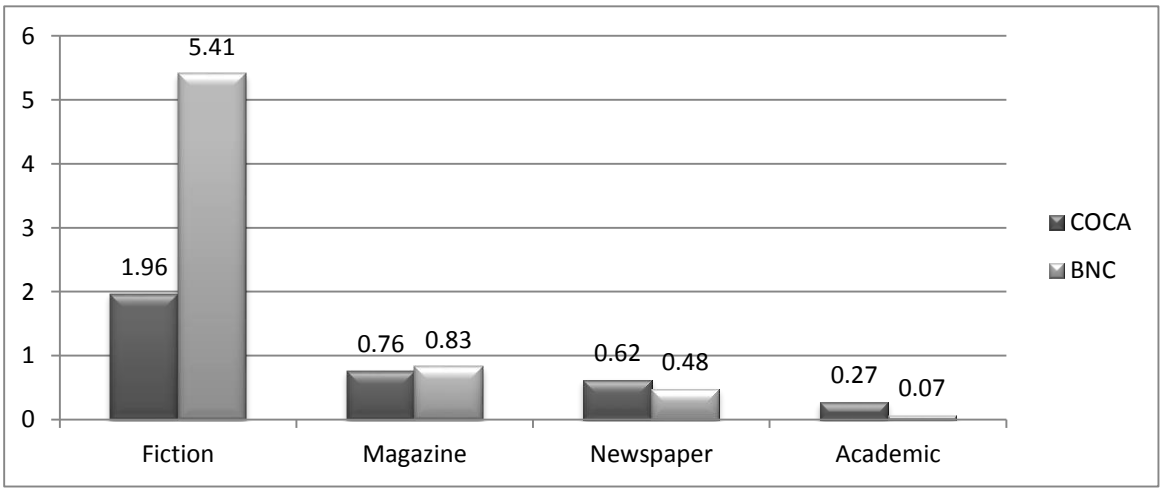
interposed between particle *a* and the node word, but I was going to fix this later, by using a wildcard in a separate search. Surprisingly, the frequencies in the written material are much higher than the ones in the spoken material in both corpora. In COCA, the normalized figure in the written material is 0.9 compared to only 0.13 in the spoken one, while in the BNC, the pattern is roughly the same, with a 1.15 normalized frequency compared to 0.10, in the spoken material (see Figure 33).

Figure 33. Frequencies per million words of *lotta* in COCA and BNC in spoken vs. written material



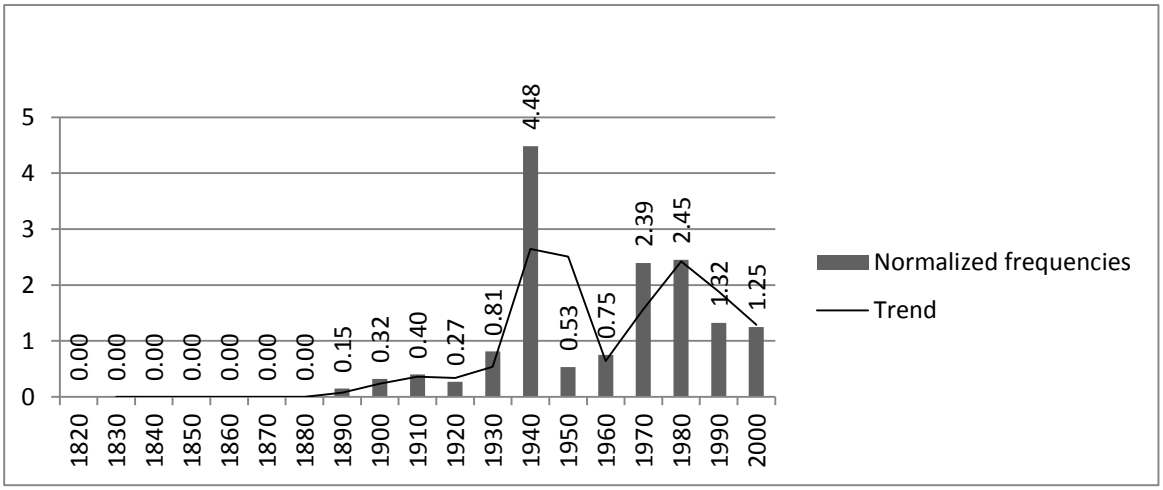
In terms of genres, the bulk of occurrences belongs to fiction in both corpora, but with a major difference between them: 5.41 for the BNC and only 1.96 in COCA, which corroborates the British 'fatherhood' of the term (Figure 34).

Figure 34. Frequencies per million words of *lotta* in the written material of COCA and BNC, divided by genre



Next comes the magazine category, with a frequency per million words of 0.76 in COCA and 0.83 in the BNC, in this case the difference between the two corpora being minor. The query returned a normalized frequency of 0.62 for the newspaper category in COCA and 0.48 in the BNC. At the bottom of the chart lies the academic genre, with a frequency per million words of 0.27 in COCA, and 0.07 in the BNC. A diachronic look at *lotta* in COHA (Figure 35) reveals a fuzzy evolution, with steep ups and downs at various times, and an unusual boom in the 1940s decade. The latest trend according to COHA is that it has decreased since the 1980s.

Figure 35. Diachronic view of *lotta* in COHA



However, the trend of the same term in Google Books British English is clearly increasing (Figure 36). The same marked increase is visible in Google Books American English, which suggests that the use of the term is on the rise in both varieties of English (Figure 37).

Figure 36. Diachronic view of *lotta* in Google Books: British English

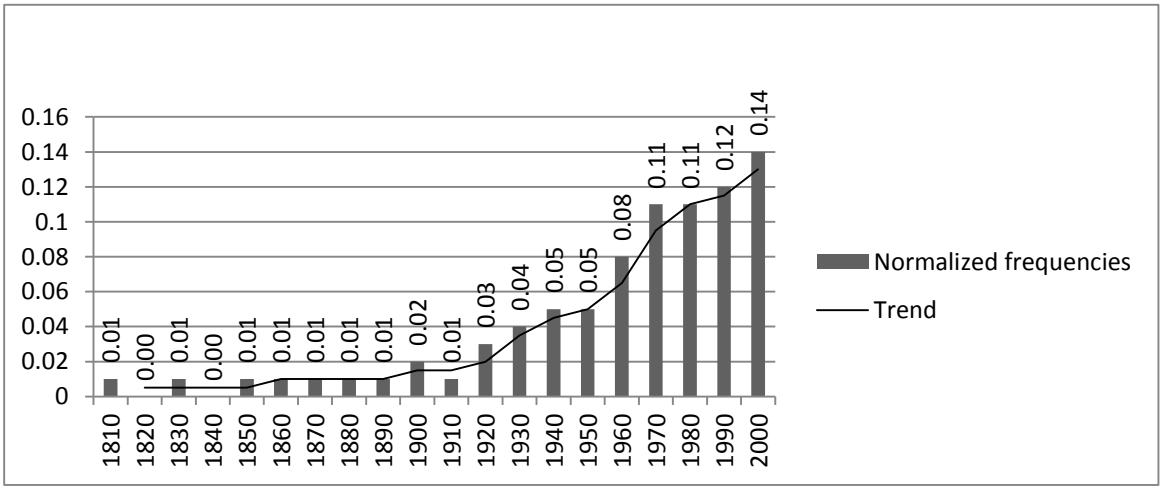
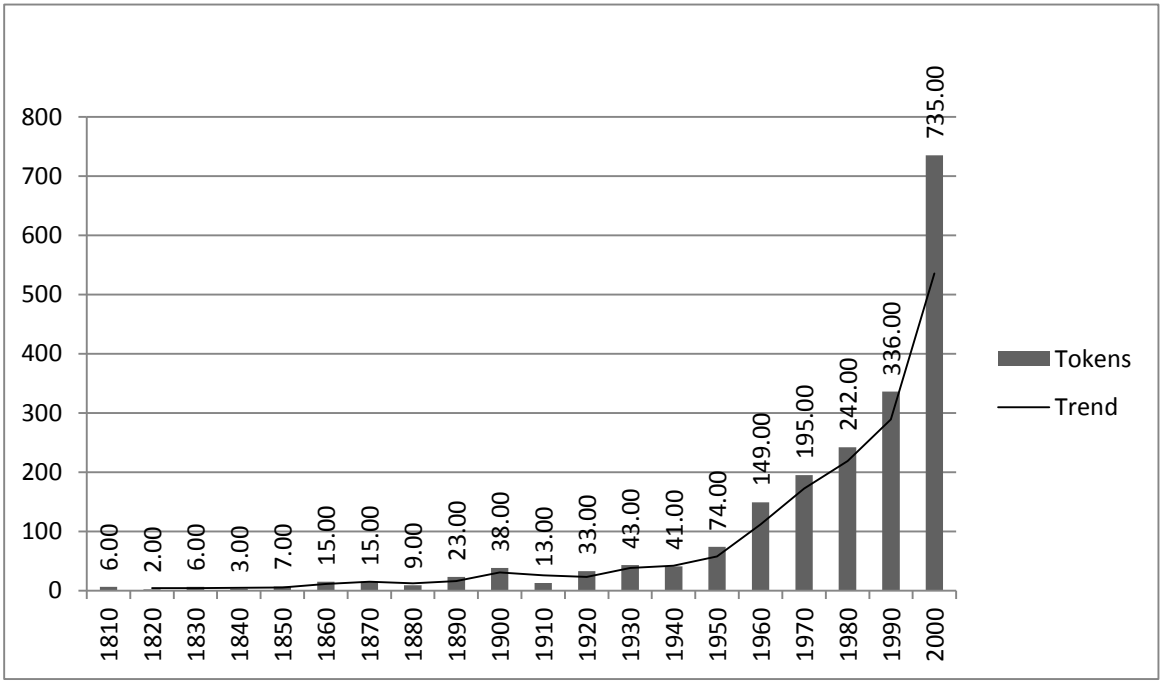


Figure 37. Diachronic view of *lotta* in Google Books: American English (Tokens)



3.5. Conclusions on Contractions

Verb contractions are quite often ignored in linguistics, but a closer look at them might reveal interesting aspects of their usage in English. As a general tendency, the use of verb contractions seems to be spreading in English. In terms of text types, the genre of fiction seems to be the most open to embracing contracted forms in writing. At the opposite end, academic writing is the most resistant to such forms, probably due to its conservatism and high formality. All standard verb contractions, with one exception, have seen a steady increase in American English during, roughly, the last two centuries. The only exception is the contracted form of *will / shall*, which has seen a decline up to the present, after a period of growth which lasted until 1910. The other contraction with a somewhat different evolution from the others is that of the verbs *would* and *had*, which had seen a decreasing trend from 1820 to 1850, but after that low point, has undergone a consistent growth up to the present day. The category of negative contractions has also seen a steady increase in American English since 1820. Like the affirmative contractions, negative contractions have the highest frequency in fiction, and the lowest in academic prose. Of the six non-standard contractions

analyzed, *gonna*, *gotta*, and *wanna* have seen a consistent increase since their emergence around the beginning of the 20th century. *Dunno* and *gimme* have had an uncertain diachronic path, with ups and downs at various times, while *ain't* has seen an abrupt decrease after a peak reached in 1910. Like the standard contractions, four of the informal items analyzed have the highest frequency in fiction, followed, in descending order, by the magazine, newspaper, and academic genres, both in American and British English. The two exceptions are *gimme* and *wanna*, which in British English have the highest frequency in the magazine category, followed by fiction, newspaper, and academic. *Ain't, dunno, gimme* and *gotta* have a higher frequency in the written material than in the spoken section of COCA, while all six non-standard contractions have higher frequencies in the spoken section than in the written material of BNC.

These findings corroborate Mair's (1997) conclusions from his own corpus-based research on the spread of the *be going to* future in the context of colloquialization:

> In the pronunciation of Standard English, speakers tolerate an increasing variety of accents. In writing, Latinate diction, long and syntactically complex "periodic" sentences formed according to the classical rhetoric are on the decline, whereas contractions, slang terms, expletives and – a lesser marker of informality – the *going-to*-future are on the rise. (Mair 1997: 1541)

4. Colloquial *Like*

One of the commonest words in English, *like* fulfills a wide range of syntactic and pragmatic functions, from standard to non-standard, some of these being still debated, as we shall see next. *Like* can be used as a noun, verb, adverb, adjective, preposition, particle, conjunction, filler, or quotative. It is the colloquial uses of *like*, be they standard informal or non-standard, that this chapter will focus on, as these fit and support the colloquialization approach. What I will try to find out, using corpus evidence, is how *like* is used as a subordinating conjunction, colloquial adverb, quotative, and filler in British and American English, and what changes have occurred, if any, in these uses of *like* over time.

4.1. A Few Theoretical Considerations

As early as in 1982, Ronald Butters observes, in *American Speech* journal, in an editor's note on Lawrence Schourup's article *Quoting with Go 'Say'*, a then brand new linguistic phenomenon consisting of the use of *be*, usually followed by *like*, to quote a speaker's "unuttered thought, as in *And he was like "Let me say something,"* or *I thought I was going to drown and I was (like) "Let me live, Lord!"* (Butters 1982: 149). This quotative use of *like* was to become, in the subsequent decades, extremely widespread among speakers of both British and American English, and will be closely investigated later on in the present chapter.

In addition to its quotative function, *like* can also be used as a **subordinating conjunction**, replacing the standard conjunctions *as if, as though.* Below is an example of this type of use of *like*, extracted from an academic text found in COCA:

(9) Say what you want about the broad appeal to civility, for a couple of days it really felt **like** we were one nation. (COCA)

This use of *like* is currently considered non-standard. *Collins English Dictionary Online* notes that until recently, the use of *like* for *such as* was unacceptable in formal writing, but now it has become accepted, e.g. *Billions of objects like/such as the one in this artist's rendering may exist in our galaxy alone*. The same is valid for *like* with the meaning of *in the same way.* The use of *as* was considered more appropriate in such a context, but now, both forms are considered equally acceptable (cf. *Collins English Dictionary Online*), e.g. *He plays chess as/like his father used to*.

Like can also be used as a **colloquial adverb**, in the construction BE + *like* + TO *infinitive,* meaning "to be likely to, to be on the verge of". Despite its long history in the English language, corpus evidence shows, as we shall see, a decline in the use of this construction in recent years. Below is an example of the use of this construction, extracted from COCA, dating back from 1883, and pertaining to the fiction genre:

(10) They came down upon us altogether with such momentum, that **we were like to be** carried from our feet by this novel charge of infantry and laid hors du combat, upon the ground. (COCA)

Another colloquial use of *like* which, unlike the adverbial colloquialism *like*, is quite new, and since its 'birth' has seen a boom in usage, especially among the younger generations in North America, is that of a so-called **quotative**. The role of such a usage of *like* is either to introduce direct speech, or some non-verbal communication. The direct speech may consist of a quotation in the form of a sentence, or simply an interjection, e.g. *I was like, "wow".* The non-verbal communication may consist of gestures, facial expressions, body movements, or making a particular sound or noise, e.g. *I was like:* [the speaker rolls his/her eyes]. Stenström *et al*. (2002: 116) note that sometimes, the BE *like* construction is used, not so much to introduce a quotation in itself, but rather to express a general feel of what was said, a meta-representation of the speaker's stance towards the content of the utterance. In (11) below, the speaker used the BE *like* construction to express his astonishment about the fact that "they" are in France, not in America, as he expected:

(11) And she goes,' Oh, they're in Paris. And I'm like Paris, California?' No, no, Paris, France. **And I 'm like ,** what? (COCA)

The same use of *like* as a quotative to express the speaker's feeling, thought or attitude about what was said is also noted by Schourup (1985):

> Speakers who use this construction claim that it prefaces not direct retrospective reports of speech, but internal speaker reactions – what the speaker had in mind to say but did not, or how the speaker felt at the time (Schourup 1985: 44)

Along the same lines as Schourup, Vandelanotte (2012) proposes an imitative interpretation of the BE *like* clauses, suggesting that the construction is used to express roughly an imitation of words, thoughts or feelings (cf. Vandelanotte 2012: 183).

Other reporting verbs, such as *say, go* or *think* are sometimes used instead of *be* in the construction, as in the example below extracted from COCA:

(12) You know, they're beautiful girls. It's just a little funny. You hear people walk by **and they go like ,**' Those are your younger sisters?' It's just that' They're your younger sisters?' (COCA)

But quotative *go* can be used either with, or without *like*:

(13) I asked if he knew Arnold Schwarzenegger, **and he goes ,** 'He's from Austria,' like I was so dumb. (COCA)

However, Andersen (2000) points out that *like* and its variants are not exactly interchangeable. While the BE *like* construction has a wide range of possible interpretations, from the speaker's thoughts to his feelings, *say* is restricted to merely introducing quotations (cf. Andersen 2000: 33). In addition to *go*, *think* and the neutral *say*, Tottie (2008: 189) mentions a further alternative to quotative *like*, namely *be all*, which, she notes, seems to be restricted to American English, and can be heard particularly among young speakers in California, as the example below, extracted from the fiction section of COCA, illustrates:

(14) "And I'm, like, 'didn't you get my messages?' **And he's all,** 'what messages?' Like his phone is broken or something." (COCA)

Schourup (1985: 38) notes that *like* is frequently used before numerical expressions, such as *like one more week.* Definition no. 2 of *like* given by the *Dictionary of American Slang – Second Supplemented Edition* (1975: 319) suggests that there is no difference in meaning between an utterance such as *It's like cold* and *it is cold.* Schourup contradicts this view, and demonstrates that there is a noticeable semantic difference between the constructions with and the ones without *like*. To support his demonstration, he gives the example of a conversation between two individuals, in which one of the interlocutors asks the other how tall he is. A reply of the type "I'm like six feet tall" from the other interlocutor would probably lead the first to ask for a more exact response (cf. Schourup 1985: 39). It can be inferred then that *like* is not always entirely functionless. Schourup (1985: 38) notes that the situations described by using such expressions are imprecise, only approximated, as in the following example extracted from the fiction section of COCA:

(15) Chase was scared and asked how much money Mom owed (but to whom? and why?) and Carly said she thought **it was like maybe** two hundred thousand dollars but Carly was only eleven that summer and not very good with numbers so it could have been much less. (COCA)

In certain dialects, *like* is often used in the end of a sentence, as parenthetic filler, meaning "as it were, so to speak" (cf. *Collins English Dictionary Online*), as in the example below retrieved from the fiction section of COCA:

(16) Bad guys couldn't just skip over a border and disappear **, like .** (COCA)

The use of *like* as a **filler** is old and common in all varieties of English. A filler is a sound, word or expression intentionally inserted in a conversation by a speaker, in order to create a delay or hesitation, and thus to gain some time to think. Some examples of words and expressions used as fillers are *well, I mean, actually, you know, let me think* (cf. the *Longman Dictionary of Language Teaching & Applied Linguistics* 2010: 220). Among the most common filler sounds in English are *uh* /ə/and *um* /əm/. Here are two famous examples from R. L. Stevenson's 1886 novel *Kidnapped*:

(17) 'Ou, ay', says the man; 'there's the laird, to be sure, if it's him you're wanting. What'll **like** be your business, mannie?' (Stevenson 1957: 6)

(18) 'Eh, man,' cried my uncle, scrambling to his feet, 'give me a meenit! What's **like** wrong with ye? I'm just a plain man and nae dancing-master; (Stevenson 1957: 191)

Another equally famous example is found in Anthony Burgess's novel *A Clockwork Orange*, where *like* is abundantly used as a discourse marker, as part of the narrator's Nadsat – the colorful fictional slang used by the teenage characters of the novel:

(19) This devochka who was **like** Pete's wife (impossible impossible) giggled again and said to Pete: 'Did you used to talk like that too?' 'Well', said Pete, and he **like** smiled… 'Well', I **like** gaped still. (Burgess 1986: 146)

4.2.Corpus Findings

Having completed the brief description of the uses of *like* which are of interest for the goal of this paper, we may now proceed with the analysis of the empirical data retrieved from the corpora.

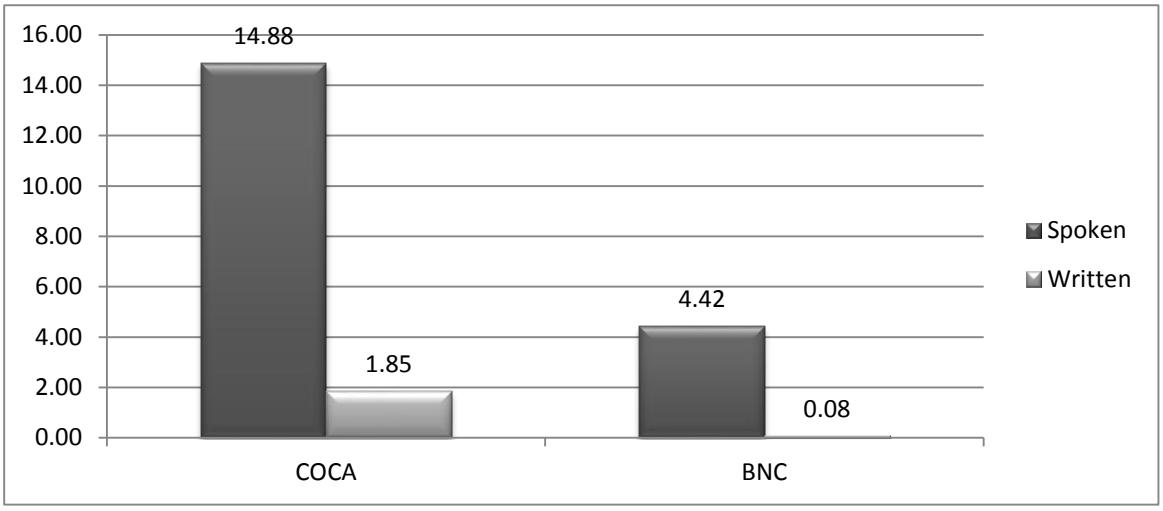
4.2.1. Quotative *Like*

Of the informal and non-standard uses of *like* which will be analyzed, quotative *like* will receive the closest attention, since, based on Buchstaller and Van Alphen (2012: XI), as well as on my own observations, it is expected to show a considerable growth in the last two decades. For reading convenience, I listed all the search strings used in this chapter in Table 12. Accordingly, for quotative *like* I used the following formula: [c*] [p*] [be] like ‚|'. I chose to include a preceding conjunction in the search string, not so much because this is quite often the case in common speech, but especially to prevent constructions which are not relevant for the quotative function of *like* from appearing in the results. The example below, retrieved from the fiction section of COCA, shows how the query results are affected by the lack of a conjunction in the search string:

(20) Father, you don't know what **it** **'s** **like** **,** really. All alone here. Always alone here. I need you so bad; (COCA)

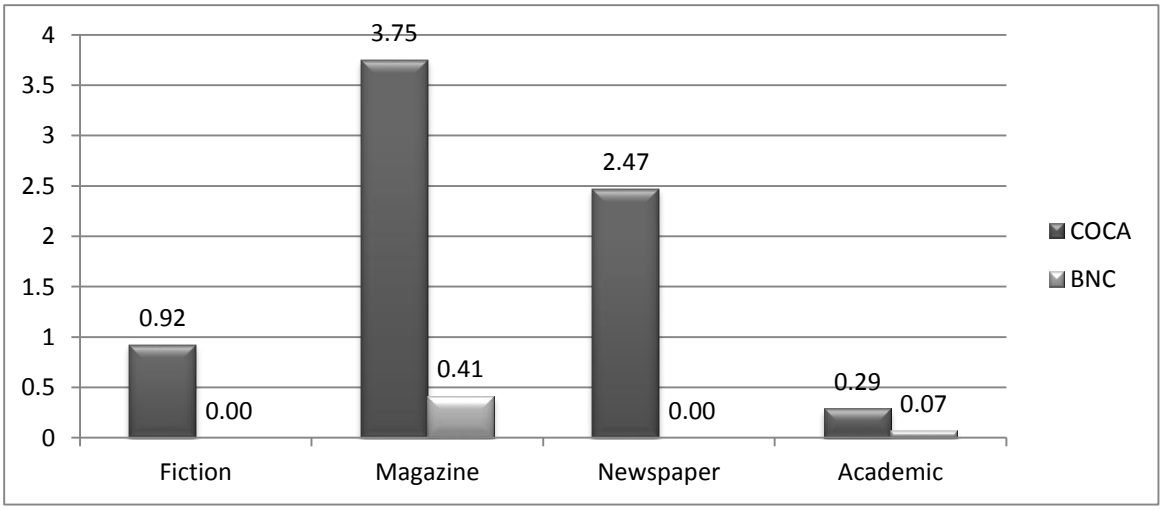
By adding a conjunction in the beginning of the search string, the accuracy of the data will not be affected by such instances of *like*. Of course, by using this particular search string, I will miss the matches which lack a preceding conjunction but where *like* is still used with its quotative meaning. As Schourup (1985:61) notes, quotative *like* is chiefly used in conversation, and this fact is illustrated in the corpora, too. There is a significant difference between the uses of quotative *like* in speech and writing, in favor of the spoken language, in both varieties of English. Thus, a majority of the occurrences of the BE *like* construction is found in the spoken section of COCA, namely a frequency per million words of 14.88, while in the written sections combined, BE *like* is poorly represented, with a normalized frequency of only 1.85. The same significant difference between speech and writing applies to British English, too, with a normalized frequency of 4.42 in the spoken material of the BNC, and only 0.08 in the written sections as a whole (see Figure 38).

Figure 38. Frequencies per million words of quotative *like* in COCA and BNC in spoken vs. written material



Significant is also the difference between the uses of BE *like* in the two varieties of English. The spoken frequency in COCA is roughly three times higher than the one in the BNC, while in writing, the difference is even higher in the American variety of English, namely 23 times higher. A look into the genres reveals further interesting facts. In American English, the Magazine category leads the way, with a normalized frequency of 3.75, followed by Newspaper, with 2.47, Fiction, with 0.92, and Academic writing, with 0.29 normalized frequencies. In the BNC, there are no matching strings in Fiction and Newspaper, while in the Magazine and Academic categories the query returned normalized frequencies of 0.41 and 0.07, respectively (Figure 39).

Figure 39. Normalized frequencies of quotative *like* in the written material of COCA and BNC, divided by genre



The substantial difference in the use of quotative *like* between British and American English in favor of the latter, suggests that this new way of introducing direct speech is preferred by the speakers American English. This corroborates the findings of Stenström *et al*. (2002: 117), who used data from the Bergen Corpus of London Teenage Language (COLT), compiled in 1993 from recorded conversations of teenagers in the London area, which suggest that the construction was not so widely spread in British English as in American English at that time. Only 34 of the 94 recorded instances of BE *like* were related to the quotational function. However, this figure constitutes 36.2% - the highest percentage of the total. The poor representation of the construction in the BNC can also be explained by the outdated status of the British corpus, considering the novelty of the expression. In all genres, not only in the Academic category, the expression appears as a quotation, sometimes within another quotation, and it probably never represents the language of the author. For example, below are two instances of quotative *like*¸ the first from the Academic section of COCA, and the other one from Fiction:

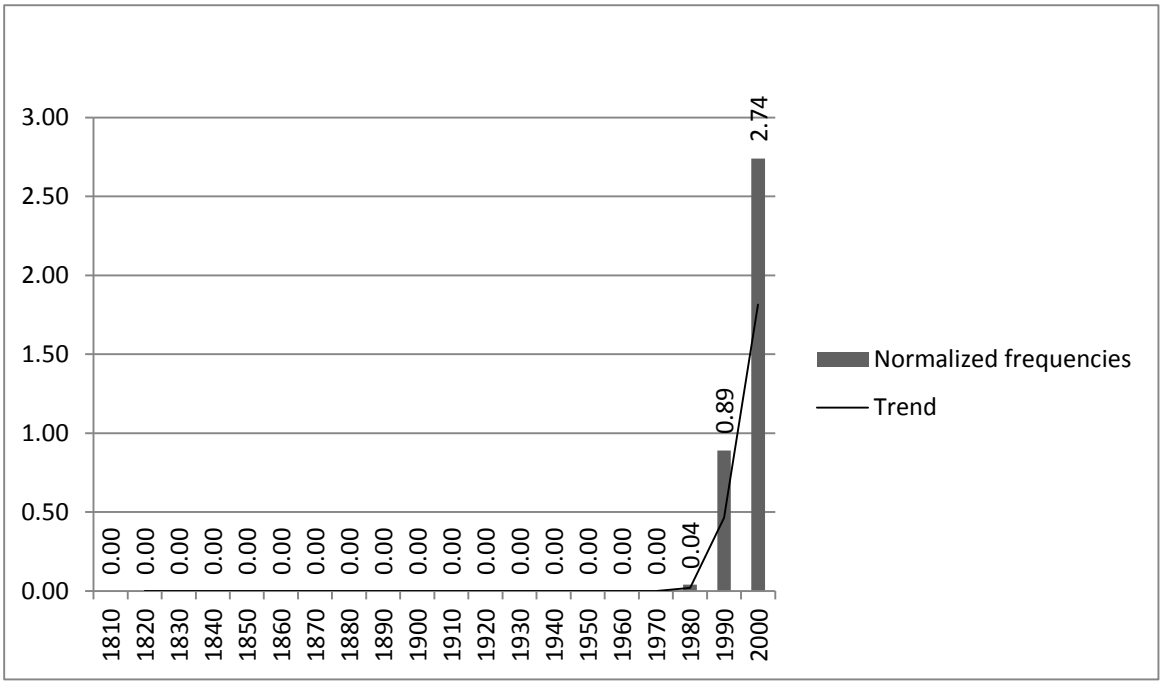
(21) And it was my first time **and I was like ,** oh, my God, it's gon na be so much fun, you know. (COCA)

(22) **And she 's like ,**' Sex sells. Go put on a tank top.' (COCA)

Although the source of quotative *like* in (21) is an academic journal, namely the *Journal of Drug Issues*, it is quite obvious that the expression was used within a quotation.

A diachronic look at the BE *like* construction in COHA (see Figure 40) reveals primarily the novelty of the expression, and its boom in the last two decades. If in the 1980s the frequency per million words was only 0.04, in the 2000s it suddenly climbed to 2.74, an increase in frequency of 68.5 times.

Figure 40. Diachronic view of quotative *like* in COHA



The same applies to the Google Books American English Corpus: if for the 1980s the query returned only 3 tokens, in the 2000s there were 143 (Figure 41). Even if it is a typical American expression, a similar marked increase is suggested by the Google Books British English Corpus. In this case, the number of tokens has doubled in the last decades, from 22 in the 1980s to 43 in the 2000s (see Figure 42).

Figure 41. Diachronic view of quotative *like* in Google Books American English

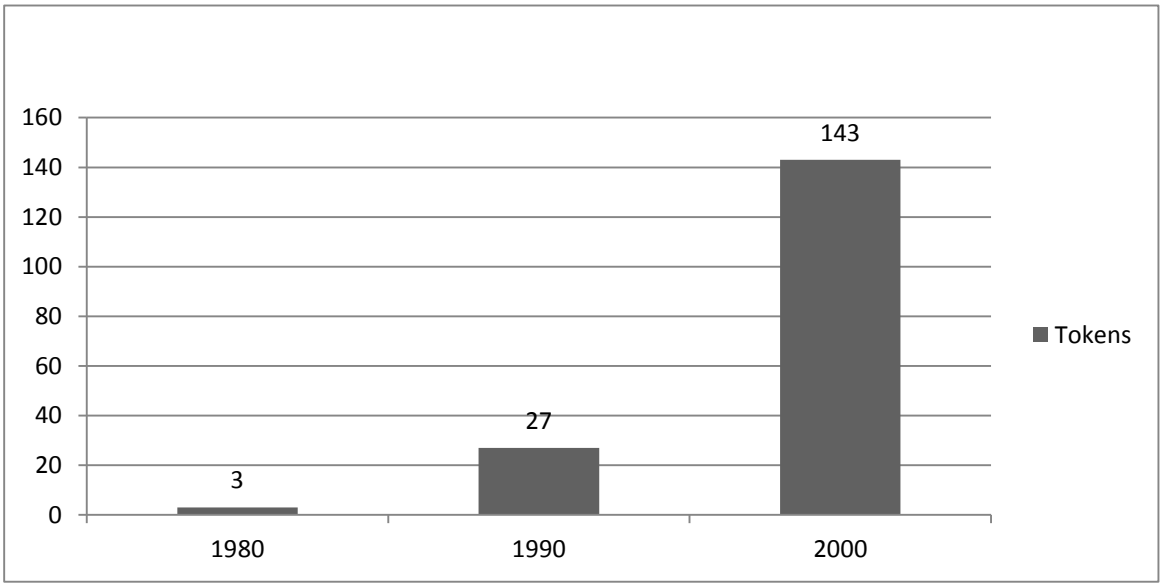
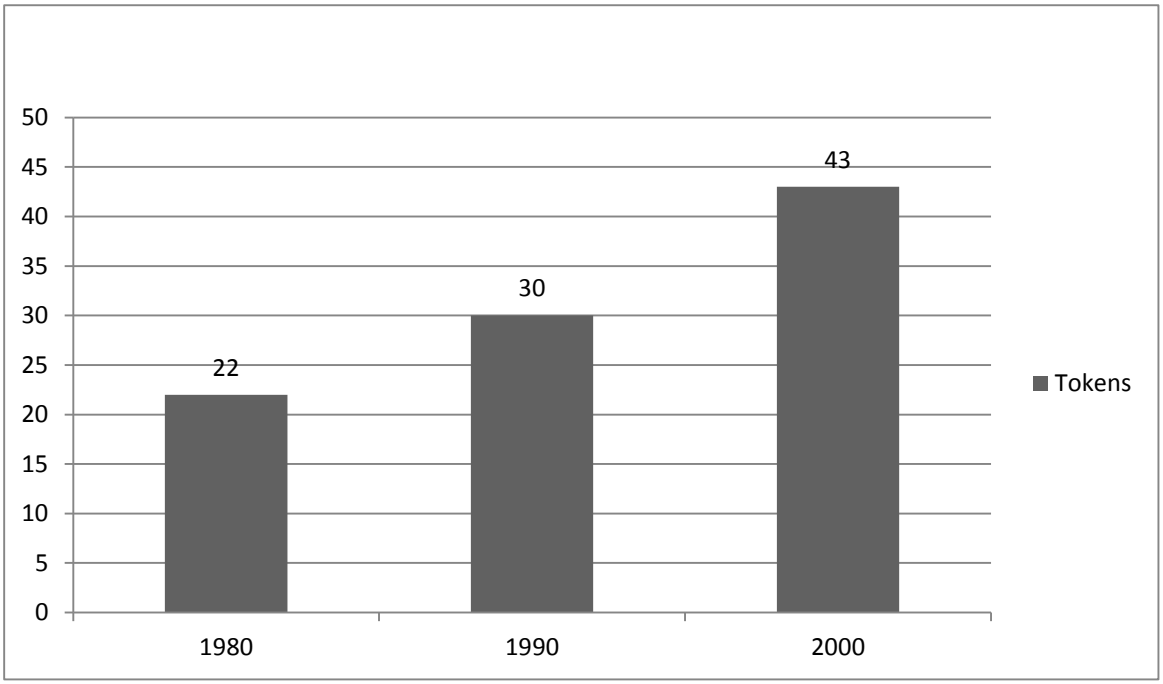
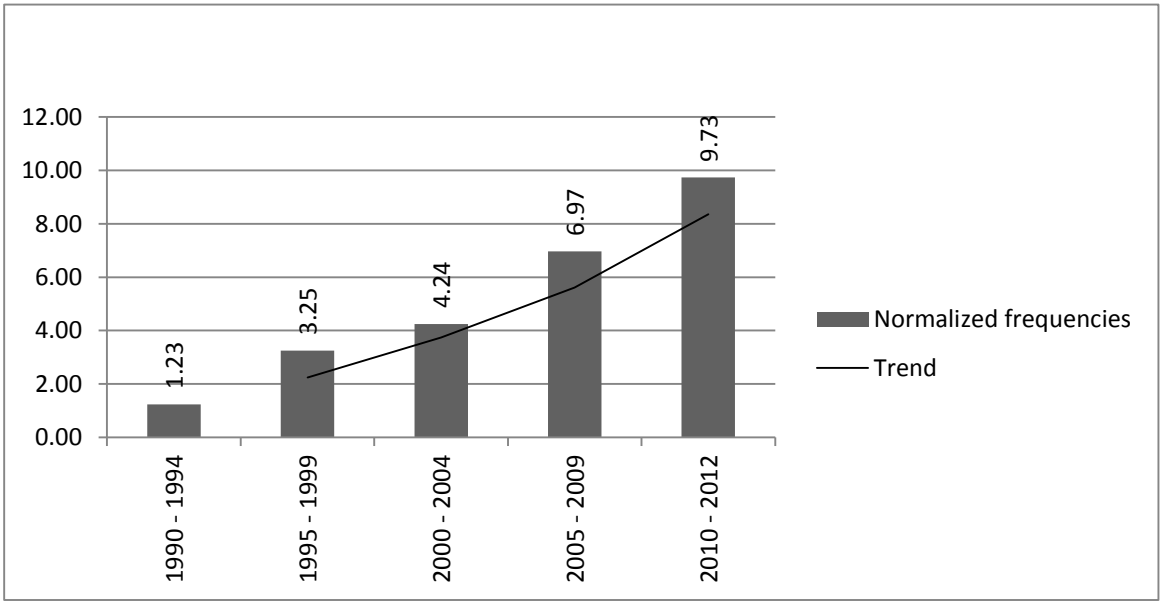


Figure 42. Diachronic view of quotative *like* in Google Books British English



Maybe the best illustration of the current ascending trend of the BE *like* construction can be seen in a diachronic query in COCA, for the last two decades, from 1990 to 2012. Although such a time period is usually considered too short for a linguistic phenomenon to be noticed and recognized as an ongoing change, in this particular case one can still see the marked growth from a frequency per million words of 1.23 in the period between 1990 and 1994 to 9.73 in the period between 2010 and 2012 (Figure 43).

Figure 43. Diachronic view of quotative *like* in COCA (1990 – 2012)



Therefore, all signals seem to indicate that quotative *like* is becoming more and more popular, at the expense of other ways of introducing a quotation, especially in American English. This fact is also confirmed by Buchstaller and Van Alphen (2012: XI), who, in turn, cite William Labov's opinion given before the opening of the 17th Sociolinguistics Symposium in Amsterdam, regarding any recent notable linguistic changes in English:

> … He mentioned "a new form in English that popped up fifteen or twenty years ago to introduce direct reported speech. Quite suddenly the old way of using *I say* or *He goes* was replaced by *I'm like* and *he's like*. Labov pointed out that this form "has penetrated as far as Australia and that it has by now become the way to start a quotation. This change went at lightning speed." (Buchstaller and Van Alphen 2012: XI).

4.2.2. *Like* as Adverb

Going further, I will now analyze another unconventional use of *like*, namely that of a colloquial adverb. As laid out in Table 12, this form consists of BE + *like* + TO *infinitive*, and the search string that I used to retrieve matches for this construction from the corpora is I|you|he|she|we|they [be] like to [v?i*]. I opted for the enumeration of some pronouns in the search string instead of using the generic pronoun tag [p*] in order to prevent the neuter personal pronoun *it* from appearing in the search results, which would affect the data accuracy through irrelevant constructions, like the one in the example below, retrieved from the Newspaper section of COCA:

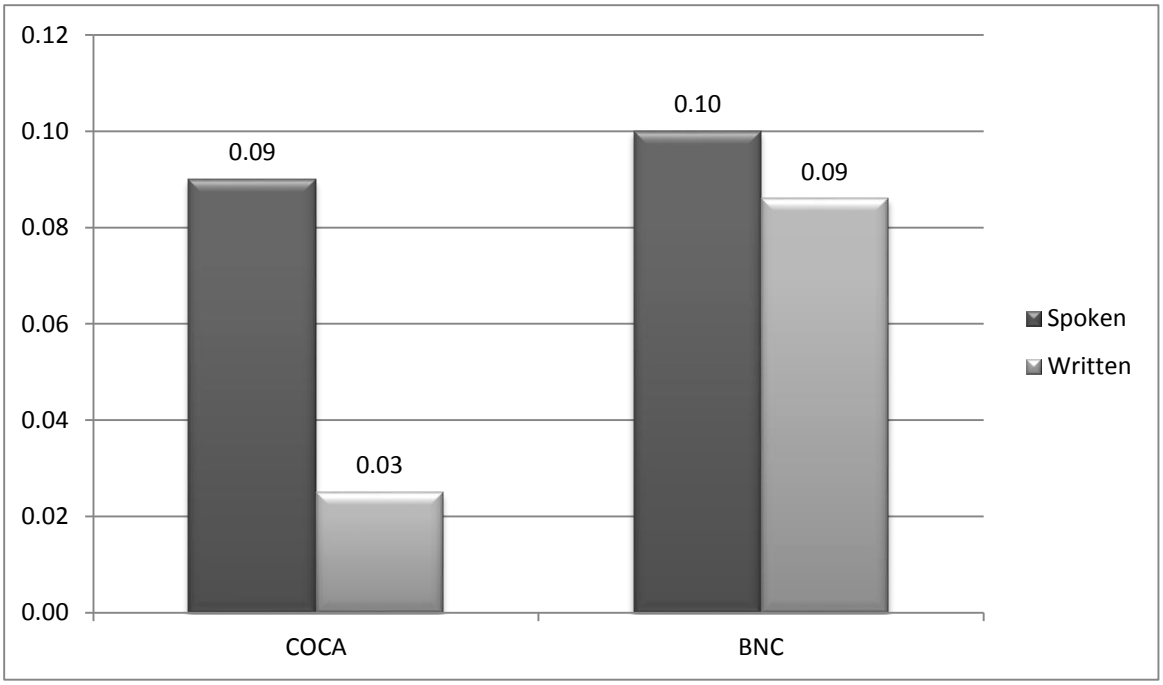
(23) " The best thing about President Obama is he knows what **it 's like to be** poor and worried, and not like Romney, who was born with a silver spoon, " says Laura Marantette, 52, of Gooding, Idaho. (COCA)

By specifying the desired pronouns in the matching strings, such irrelevant constructions will no longer appear in the query results, and the data accuracy will be maintained. See for instance the following utterance extracted from the Fiction section of COCA:

(24) That damn Cooper just got a hold of it and God damn, I thought **I was like to faint**. (COCA)

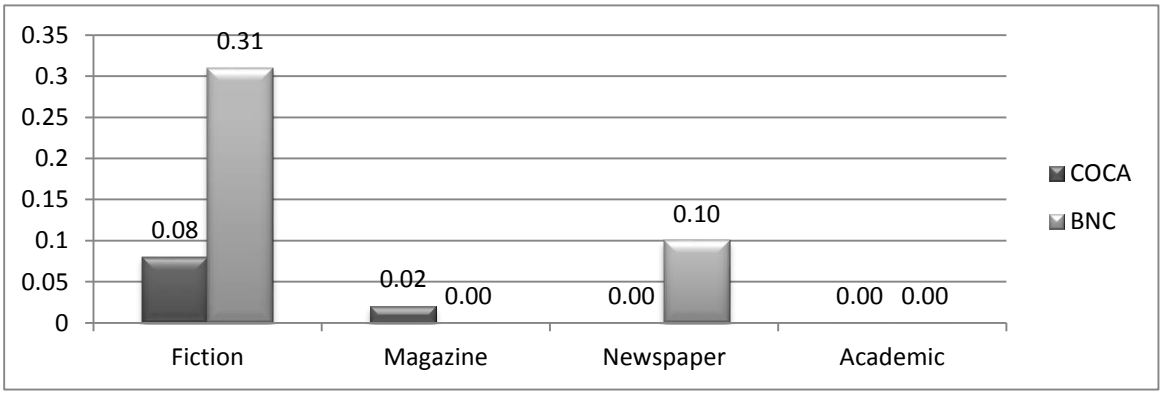
As a colloquial adverb, *like* returned a normalized frequency of 0.09 in the spoken material of COCA, and 0.025 in the written sections combined (see Figure 44). In the BNC, the frequency per million words of the spoken material is 0.10; while in the written sections it is 0.086. Surprisingly, the usage is slightly higher in the BNC than in COCA, yet not very significant in either of them.

Figure 44. Normalized frequencies of *like* as colloquial adverb in COCA and BNC in spoken vs. written material



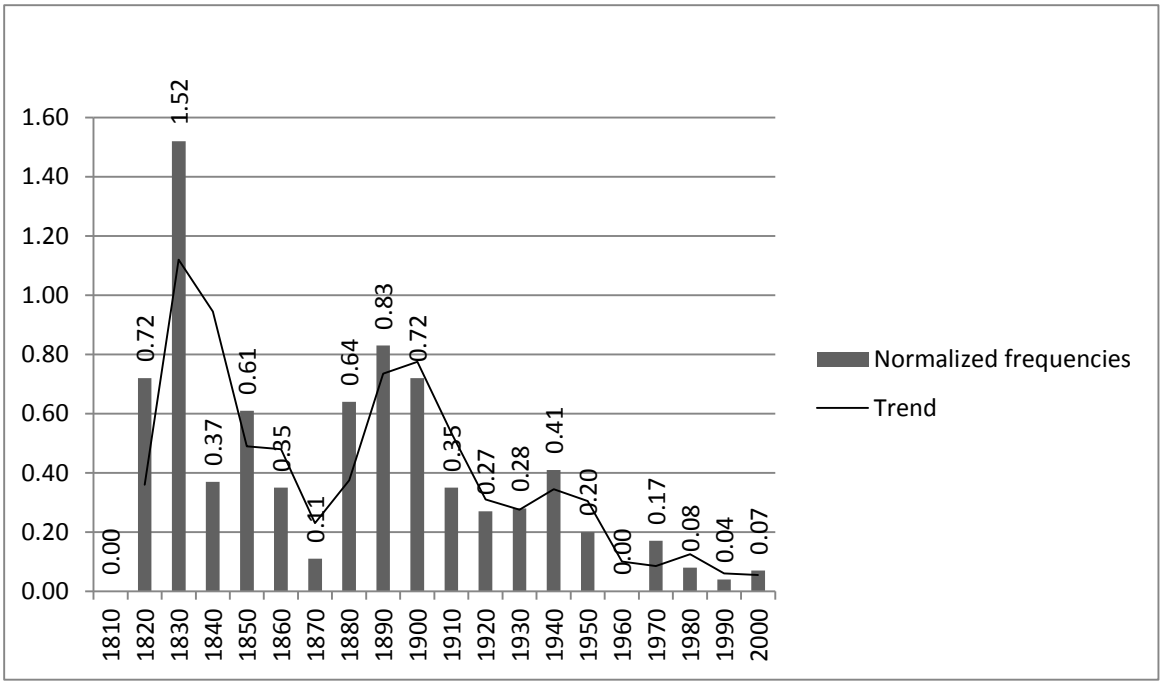
In terms of genre, COCA returned a 0.08 normalized frequency in Fiction, followed by the Magazine section, with a frequency per million words of 0.02, and no matching strings in the Newspaper and Academic categories (Figure 45).

Figure 45. Normalized frequencies of *like* as colloquial adverb in the written material of COCA and BNC, divided by genre



In the BNC, the only two genres which returned results are Fiction, with a normalized frequency of 0.31, and Newspaper, with a frequency per million words of 0.10. No matches were found in the Magazine and Academic genres. The diachronic picture of *like* as adverb in COHA suggests the decline of this use at least in American English. As Figure 46 shows, the adverbial use dropped from its peak in the 1830s, from 1.52 to 0.07 in the 2000s.

Figure 46. Diachronic view of *like* as colloquial adverb in COHA



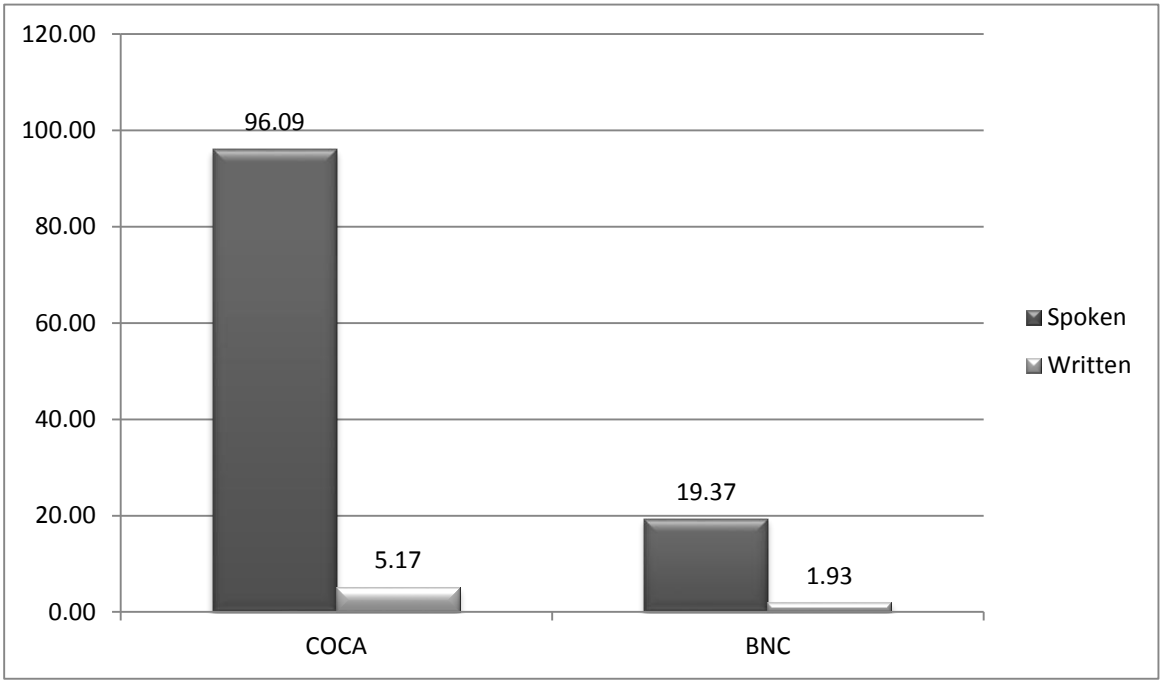
4.2.3. *Like* as Filler

Unlike the adverb *like*, the use of the filler *like* is widespread in all varieties of English, and this is reflected in the corpora, too. There is no particular search string for this use of *like.* By simply placing the word between commas, one can get the desired results from the corpora. The example below, retrieved from the Fiction section of COCA, shows how *like* is used as a filler, intentionally inserted in the utterance by the speaker, to create a break in the speech flow and to approximate the mentioned distance:

(25) I could feel the magic coming off the group when they were still **, like ,** twenty yards away. (COCA)

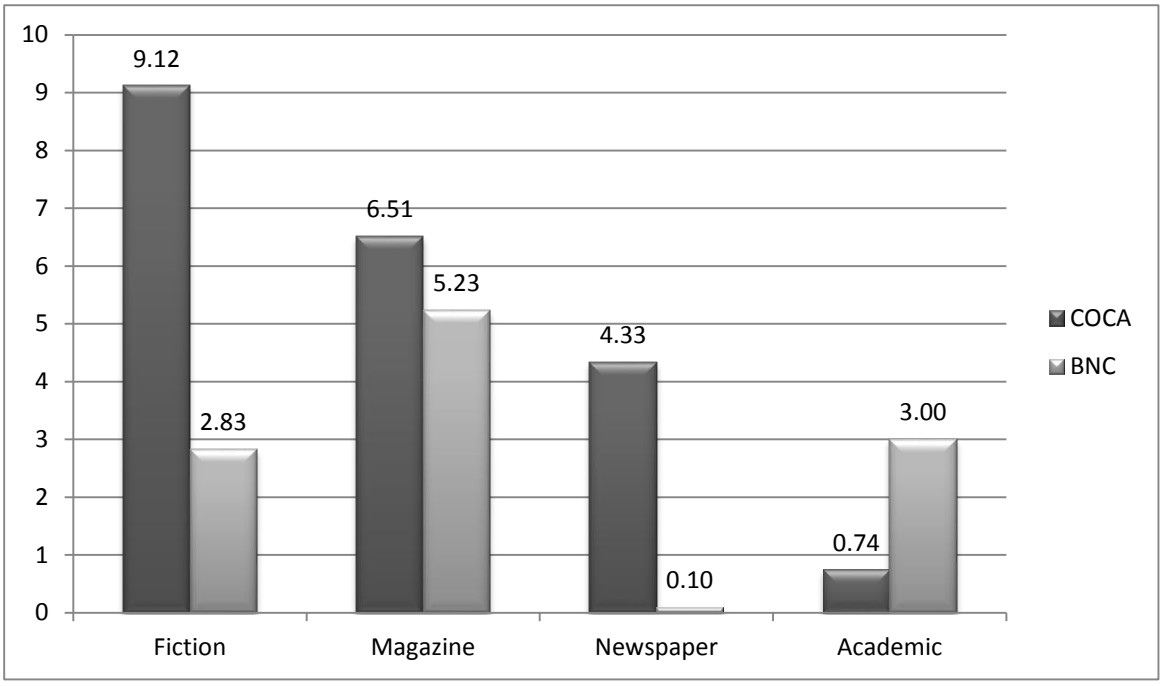
In Figure 47, the extensive use of *like* as a filler in speech in American English can be seen, with a normalized frequency of 96.09. Compared to this, the written genres of COCA give a frequency of 5.17 per million words combined. One can also notice the significant difference in this use of *like* between British and American English, in favor of the latter. A normalized frequency of only 19.37 was found in the spoken section of the BNC, and of 1.93 in the written material as a whole.

Figure 47. Normalized frequencies of *like* as a filler in COCA and BNC in spoken vs. written material



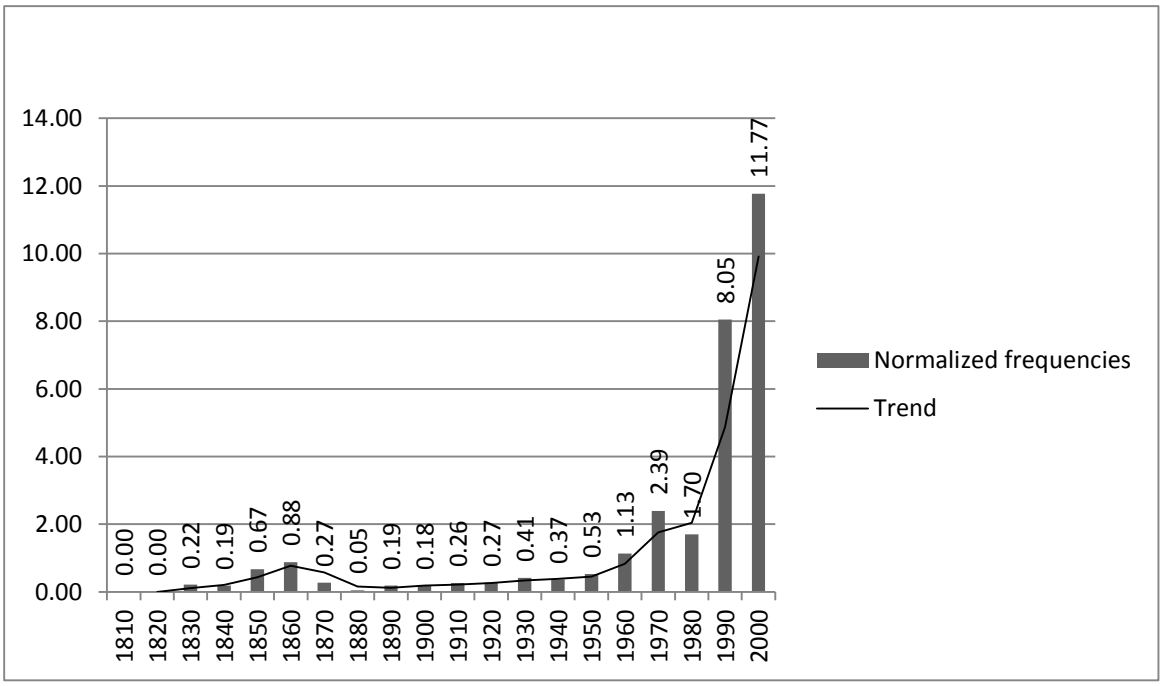
By genre, Fiction holds the leading position in COCA, with a normalized value of 9.12, followed by Magazines, with 6.51, Newspaper, with 4.33, and Academic, with 0.74 (see Figure 48). In the BNC, the situation is different, the Magazine category holding the highest frequency, surprisingly followed by the Academic genre, with a frequency per million words of 3.00, Fiction, with 2.83, and Newspaper, with only 0.10. The apparently high frequency of the filler *like* in the Academic category of the BNC is due to only 46 tokens (the total number of occurrences), which is not much for a corpus the size of the BNC, and moreover, this high frequency, compared to other genres in the corpus or to American English, is probably due to the arbitrary, uneven composition of the corpus.

Figure 48. Normalized frequencies of *like* as a filler in the written material of COCA and BNC, divided by genre



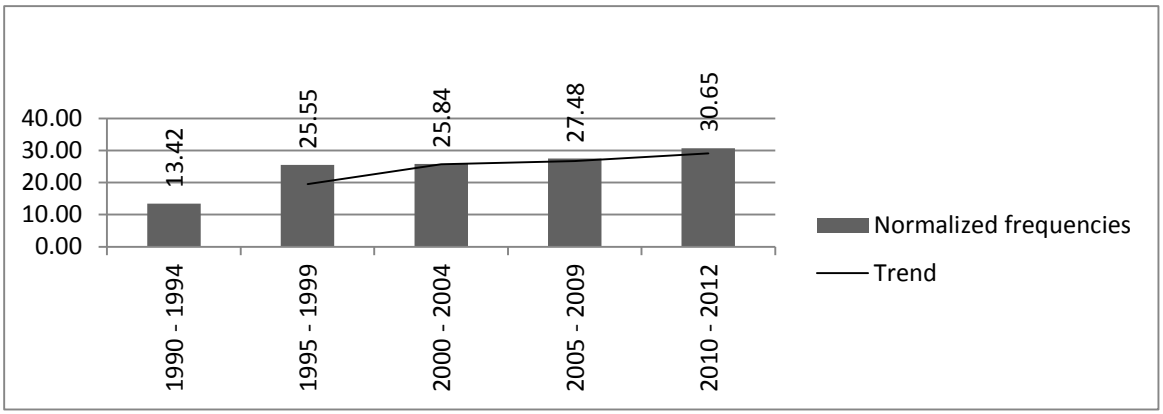
Diachronically, the use of *like* as a filler in American English seems to have seen a marked increasing trend since the 1980s, from 1.70 normalized frequencies to a stunning 11.77 in the 2000s, as Figure 49 shows.

Figure 49. Diachronic view of *like* as a filler in COHA



A similar growth can also be perceived even for a shorter and more recent period of time, as a diachronic query in COCA proves (see Figure 50). This usage of *like* has doubled, from a frequency of 13.42 per million words in the period between 1990 and 1994, to 30.65 in the period 2010 – 2012.

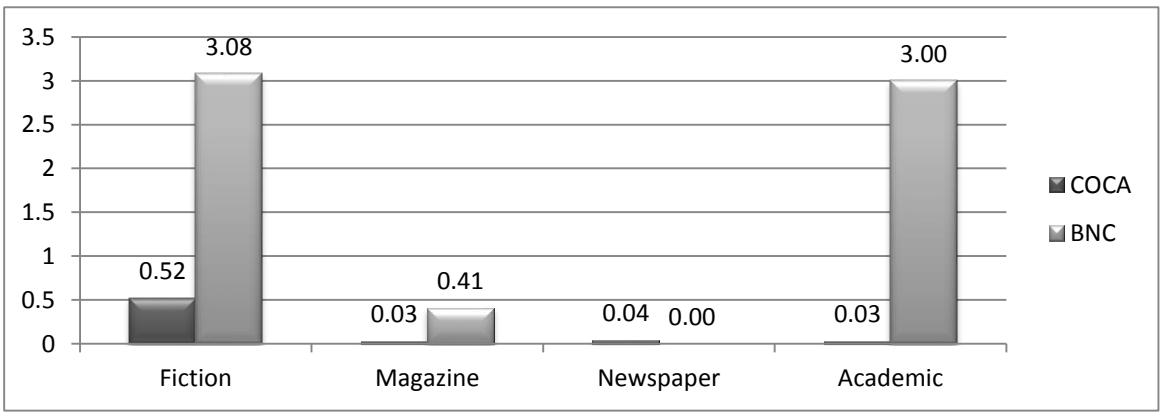
Figure 50. Diachronic view of *like* as a filler in COCA (1990 – 2012)



A particular positioning of the filler *like* is at the end of the sentence, as a replacement for *you know*, with the meaning *as it were, so to speak.* In order to retrieve instances of the sentence-final filler *like* from the corpora, the word must be placed between a comma and a full stop. The considerably higher frequencies in the Fiction, Magazine, and Academic genres of the BNC than in COCA demonstrate that this use of *like* is rather a feature of British English than of the English spoken in the United States (Figure 51). In (26) below *like* is used as a filler, placed at the end of the sentence, in an academic paper excerpt from the BNC:

(26) Annie was real clever – even though she was only a little kid **, like.** (BNC)

Figure 51. Normalized frequencies of the filler *like* placed at the end of the utterance in the written material of COCA and BNC, divided by genre



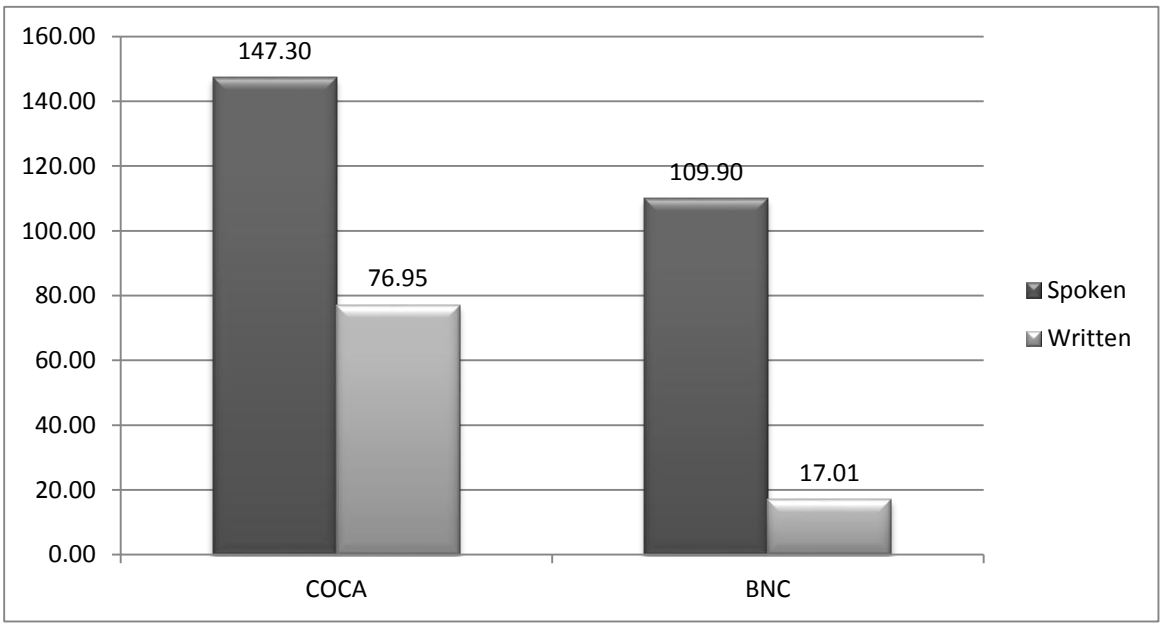
4.2.4. *Like* as Subordinating Conjunction

Despite the current half non-standard status of *like* as a subordinating conjunction, i.e. non-standard with the meaning *as if*, standard with the meaning *in the same way as*, deemed undesirable in formal writing until recently, this usage has seen the same growth as *like* as a filler has seen in the last twenty years. No particular search string is required to retrieve matches of this use of *like* from the corpora, except, of course, its tagging for the desired part of speech, that is, like.[cs*]. In (27) such an instance of *like* is used as a substitute for *as if*, in an archeological academic paper:

(27) Footprints are still visible in the fields, and within the collapsed walls of a courtyard, intact millstones and ceramic food steamers look **like** they are waiting to be used. (COCA)

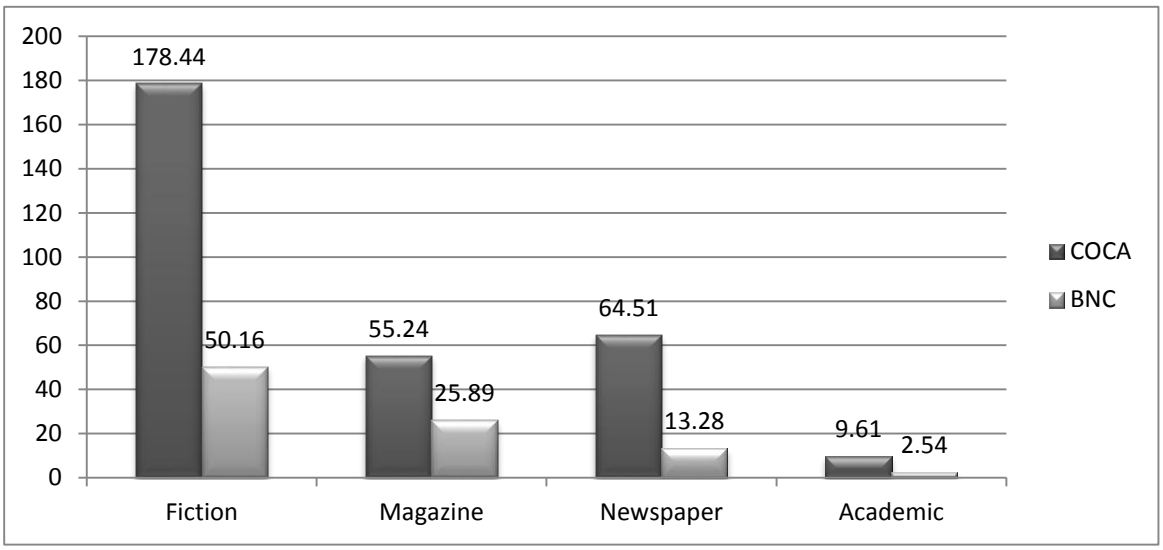
In American English, the conjunction *like* stands out in speech, with a normalized frequency of 147.30, while the written material has also reached a high level of usage, with a frequency of 76.95 per million words (see Figure 52). In British English, the query returned a normalized frequency of 109.90, whereas in the written genres of the BNC as a whole, the frequency per million words is 17.01.

Figure 52. Normalized frequencies of *like* as subordinating conjunction in COCA and BNC in spoken vs. written material



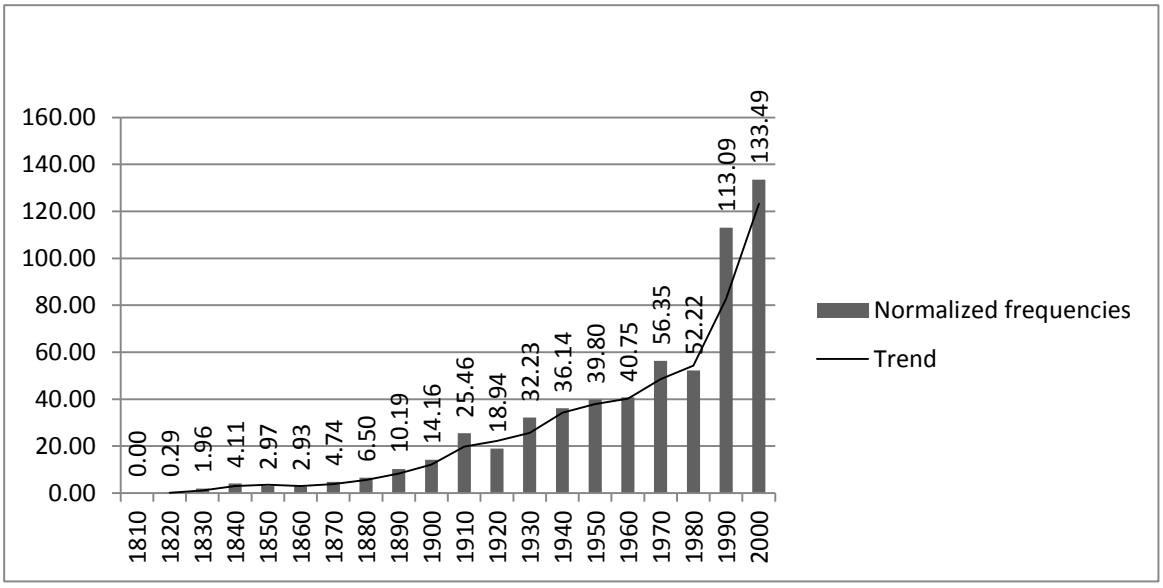
As far as genres are concerned, Fiction leads the way in American English, with a normalized frequency of 178.44, followed by Newspaper, with 64.51, Magazine with 55.24, and Academic writing with 9.61 (Figure 53). In British English things are slightly different, in the sense that the Fiction genre tops the frequency chart, with a normalized value of 50.16, followed by Magazine, with 25.89, Newspaper, with 13.28, and Academic, with 2.54.

Figure 53. Normalized frequencies of *like* as subordinating conjunction in the written material of COCA and BNC, divided by genre



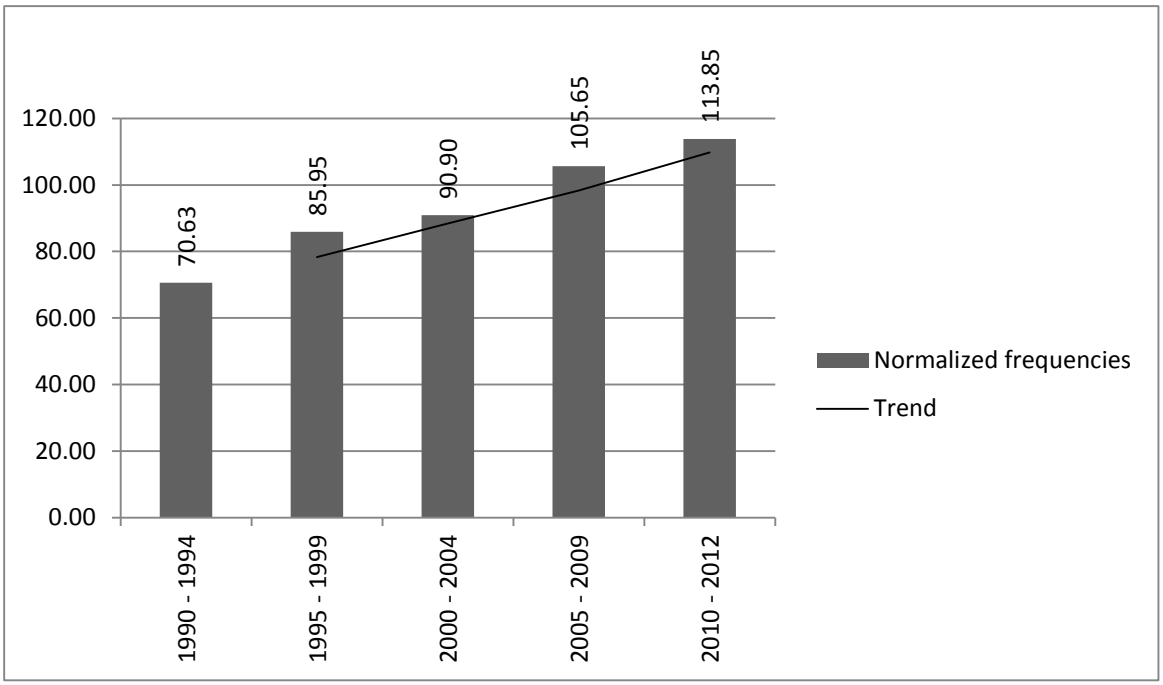
From a diachronic point of view, the conjunction *like* has seen a continuous and substantial growth since the 1820s, from as low as 0.29 to 133.49 in the 2000s (see Figure 54).

Figure 54. Diachronic view of *like* as subordinating conjunction *in* COHA



Even a look at the conjunction in the diachronic component of COCA, which spans over a much shorter period of time, that is, roughly two decades, still demonstrates an obvious growth, from a frequency per million words of 70.63 in the period 1990 – 1994 to 113.85 in the period between 2010 and 2012 (Figure 55).

Figure 55. Diachronic view of *like* as subordinating conjunction in COCA (1990 – 2012)



4.3. Conclusions

Like is a pretty versatile word in English, fulfilling a wide range of syntactic and pragmatic functions in the utterance. Some of these functions are established standard forms of English grammar, while others are still controversial. Of the informal and non-standard uses of *like*, the quotational one has had the most dynamic evolution since its birth some twenty years ago. It rapidly gained tremendous popularity, especially among the teenage speakers of American English, gradually replacing the other more conventional ways of introducing direct speech in conversation with verbs such as *say* or *go*.

The use of *like* as a colloquial adverb in the form BE + *like* + TO *infinitive* is not significant either in British or American English. Moreover, in American English, this use of *like* has markedly declined over the last two hundred years.

Like as a filler, on the other hand, is widespread in both major varieties of English. Corpus evidence demonstrates the increasing use of this discourse particle in American English, especially during the last two decades. In some dialects of British English the particle is placed at the end of the utterance as an alternative to *you know*, but this practice is not so common in American English.

Like can also fulfill the function of a subordinating conjunction, as an alternative to *as, as though, as if,* or *in the same way as*. Until recently, the use of *like* as a substitute for *as, such as* and *in the same way as* was considered infelicitous, while its use instead of *as if* is still considered non-standard. Despite its controversial status, *like* used as a subordinating conjunction has seen a consistent growth in American English, in the last two hundred years, as corpus evidence shows.

Like, in all its colloquial uses, is primarily suited for conversation, and this is also illustrated by the corpus findings. In all its hypostases, *like* is much more widespread in speech than in writing. Among the written genres, the fewest occurrences of *like* can be found in academic papers, and when it does happen, it is always used within a quotation.

The uses of *like* as a quotative, filler and conjunction support the colloquialization approach of this thesis, as these uses mark a movement from formal to informal styles of communication, and the corpus investigation carried out on these functions of *like* demonstrates their substantial growth. The observed changes in the use of *like* corroborate the findings of Mair (2007: 234) and Schourup (1985: 61):

> The current efflorescence of *like* in conversation, at least among younger speakers, is not a symptom, as Newman would have it, of "the death of English", but the spread from its originally quite restricted range of occurrence of an item which in general indicates a possible loose fit between overt expression and intended meaning (Schourup 1985: 61).

5. Phrasal Verbs and Colloquialization

5.1. General Remarks

This chapter is devoted to the study of phrasal verbs in the context of colloquialization. Phrasal verbs are associated with the informal style, and therefore they are expected to be more commonly found in speech and written genres such as fiction and journalism, and to a lesser extent in texts involving a more formal style, such as academic prose. Kennedy (1967) claims that one of the reasons why speakers frequently choose to use a phrasal verb instead of its simple counterpart is a desire to strengthen or emphasize what is said, e.g. *pay* vs. *pay up, stretch* vs. *stretch out,* etc. (cf. Kennedy 1967: 33). Another reason for the popularity of verb-particle constructions invoked by Kennedy is "mere linguistic laziness":

> It is so much easier to combine a dozen of fifteen well-known prepositional-adverbs with the simpler and more familiar verbs than it is to keep in mind a special polysyllabic verb for each different concept. (Kennedy 1967: 33)

To support his claims, he gives a number of more or less suggestive examples: *bear out* for *corroborate, find out* for *discover, keep on* for *continue, look over* for *examine, use up* for *exhaust,* and *watch out* for *beware* (cf. Kennedy 1967: 33). As for their formation, Kennedy argues that many of the verb-particle constructions are created by analogy with other older, already established forms, similar in meaning. For a verb like *add up,* for instance, several alternative combinations with a similar meaning are available, e.g. *count up, figure up,* or *sum up* (cf. Kennedy 1967: 33). Kennedy argues that the rapid development of phrasal verbs is owed to the common, relatively uneducated social classes (cf. Kennedy 1967: 40). This is not necessarily true, at least nowadays, considering that the first edition of Kennedy's book is dated 1920. It is very likely that a computer geek uses a plethora of colloquial and slang expressions, but this is not to say that he is uneducated or has a common mind. Despite his colorful language, he may still be a computer genius. What I would rather infer from Kennedy's opinion is that verb-particle constructions identify themselves with colloquiality and informality, as he puts it in another statement:

> On the other hand, the words that are being displaced, such as, for example, *collapse* (break down), *comprehend* (catch on), *eliminate* (cut out), *issue* (give out), etc., were introduced into the language, for the most part, through scholarly influence. Many of

> these words are directly chargeable to the artificial increase of our vocabulary by that school of 'embellishers' who flourished in the sixteenth and seventeenth centuries. Perhaps the common man is not to be blamed for avoiding the use of a vocabulary which has never really been his, and for utilizing in the expression of certain ideas his own familiar stock of words. (Kennedy 1967: 40)

Socially speaking, Kennedy suggests that the spread of phrasal verbs took place from the bottom up, eventually reaching all classes of English speakers. Therefore, the spread of verb-particle constructions into all social layers of English users, at the level of both speech and writing can be considered part of the process of colloquialization. But is there an increase in the use of phrasal verbs in present-day English? If so, what is their current status and distribution based on various criteria in the two main varieties of English? Hundt and Mair (1999) have already found an increased use of phrasal verbs, particularly of those including the particle *up*, in the journalese of the LOB-FLOB and Brown-Frown pairs of corpora, but a decline of their use in the academic genre, especially in American English. The verbs which will be mainly scrutinized in this chapter are those constructed with what is said to be the most productive particles in the English particle-verb constructions, namely *up* and *out.*

5.2. A Brief History of the English Verb-Particle Constructions

Before starting a detailed description of the verb-particle constructions in present-day English, I believe it is worth taking a quick look at the long-term historical development of these constructions.

In Old English, the particle was, in most cases, prefixed to the verb, e.g. the Modern English verb *go out* was *utgan* in Old English (cf. Gries 2003: 3). The particles could not be separated from the verbs, and the meaning of this type of complex verbs was literal, that is, directional or spatial (cf. Gries 2003: 3). Relics of Old English verbs with an inseparable prefix have survived in contemporary English in such forms as *forgive, foreshadow, outface, outnumber, overtake, overthrow, understand, undertake, withstand,* and so forth (cf. Kennedy 1967: 11). The split between the prefixed particles and the accompanying verbs took place gradually, in a two-stage process, known as the First Particle Shift (cf. Gries 2002: 4). In the first stage, the separation of the particle from the verb became possible, but its position had to be only before the verb stem. In the second step, the placement of the particle after the verb became possible also, but this could only happen with finite verbs. At this stage, non-finite verbs could not be

followed by the detached particle (cf. Gries 2003: 4). A few centuries after the First Particle Shift, around 900 A.D., the placement of particles after non-finite verb forms became also possible, this being known as the Second Particle Shift (cf. Gries 2003: 4). The two linguistic processes were completed in the Middle English period, by around 1200 – 1300 A.D., when the post-verbal position of the particle became predominant, and when verbs such as *bringan up*, *cuman up* or *giefan up* started to be used with other meanings than the literal, spatial one (cf. Gries 2003: 5). Gries notes that "until and during Shakespeare's time, phrasal verbs were frequently used in both colloquial and written language" (Gries 203: 5). He also notes that the earliest examples of particles occurring either adjacent to the verb, or following the object, can be found in texts by Alfred the Great (899) and Abbot Ælfric (ca. 1020) (cf. Gries 2003: 5). Kennedy (1967) notes that in the first 300 lines of *Beowulf*, there are no less than 25 instances of verbs with inseparable prefix, e.g. *ofteath, forgeaf, becom*, and only 5 with a separate adverbial modifier, e.g. *up ahafen, ut scufon* (cf. Kennedy 1967: 11). According to Hiltunen (1983: 6), by ca. 1250 A.D., phrasal verbs had attained their present-day form and features.

5.3. The Main Features of Present-Day Verb-Particle Constructions

Dehé (2002) notes three categories of verb-particle constructions in English: **transitive particle verbs**, e.g. *the man helped out the little girl*, **intransitive particle verbs**, e.g. *she grew up in a small town*, and **complex particle verb constructions**, e.g. *Andrew will print his teacher out a copy* (Dehé 2002: 3). The main characteristic which differentiates among transitive verb-particle constructions is the placement of the particle, either adjacent to the verb or swapped with the direct object, provided this is a noun phrase, as in the two examples below:

(28) He *wiped off* her tears.

(29) He *wiped* her tears *off*.

Dehé (2002: 3) calls these two types of constructions **continuous** and **discontinuous**, and points out that with unstressed pronouns, the swapping between the particle and direct object is obligatory, as in the following two examples:

(30) He wiped *it* off.

(31) *He wiped off *it*. (Dehé 2002: 4)

Another differentiation which can be made among verb-particle constructions is in terms of their semantic properties. Thus, Dehé (2002) distinguishes among three such categories: **compositional** or **transparent,** where each of the constituents, namely the verb and the particle, has its literal meaning, which in general has to do with direction or space; **idiomatic particle verb constructions**, in which the constituents form a semantic unit whose meaning cannot be inferred from their individual meanings; and **aspectual particle verb constructions**, where the particle is also semantically transparent, and typically marks the action telic, that is, the situation is goal oriented and regarded as complete. Below are three examples, one of each of the categories described above, taken from Dehé (2002: 6):

(32) James *carried up* the suitcase. ('up the stairs'; compositional)

(33) The baby *threw up* the meal. ('vomit'; idiomatic)

(34) John *ate up* the cake. (aspectual)

Thim (2012: 13) comes up with a slightly different classification, in the sense that he reunites the directional and aspectual particles under the same category of compositional constructions, and calls the idiomatic constructions non-compositional (see Table 3).

Table 3. Semantic classification of phrasal verbs (based on Thim 2012: 13)

Verb-particle combination		
Compositional		Non-compositional
Directional particle	Aspectual particle	

He argues that aspectual particles might be regarded as a subgroup of the compositional category, because both aspectual and directional particles are semantically transparent, and can combine quite freely with any verb, in improvised, fully logical constructions.

Thim also notes that the relatively free exchangeability of the compositional verb-particle combinations is dependent on the interpretation of the whole construction as 'motion through

space', where the particle indicates the direction, and the verb expresses the kind of action (cf. Thim 2012: 14):

(35) John *took / put / threw /carried* the book *away / up / in / back / out.*

Therefore, the nature of the direct object is one of the factors governing the placement of the particle, but Dehé (2002: 76) points out that particles can occur adjacent to the verb even with unstressed pronouns, provided they are focused, as in the following excerpt from William Faulkner's short story 'The Bear', found in the Fiction section of COCA:

(36) If he gets hemmed up and has to pick out somebody to run over, he will **pick out you .** (COCA)

Among the many factors proposed to account for the choice of the word order in the verb-particle constructions, one of the best known is the "length or syntactic complexity of the direct object" (Dehé 2002: 76). Thus, full noun phrases may occur either before or after the particle, while heavy objects are normally placed at the end of the construction:

(37) Sometimes they **carried up a bucket** of sand and sometimes tiles. (COCA)

(38) We **carried the coffin up** onto the porch. (COCA)

(39) **Look up** the dictionary that I gave you last week.

(40) ***Look** the dictionary that I gave you last week **up**.

However, if with directional particles the placement is free, in idiomatic constructions, the particle position is restricted to the slot next to the verb:

(41) You have said that when someone you love is sick, it's important to not **give up** hope. (COCA)

(42) ??You have said that when someone you love is sick, it's important to not **give** hope **up** .

But, if as already mentioned, the object is a pronoun, the particle-object inversion is compulsory even if the phrasal verb is idiomatic:

(43) Lying is a bad habit of mine and I should **give it up**. (COCA)

(44) *Lying is a bad habit of mine and I should **give up it** .

A search in COCA for the constructions in (42) and (44) above did not return any matches, which confirms their incorrectness.

Since, generally speaking, particles in English are morphologically identical with prepositions (e.g. *up, out, off*) and adverbs (e.g. *away, back*), what differentiates between them is the exclusive ability of the particle to occur adjacent to the verb, before the direct object (cf. Dehé 2002: 4):

(45) Nicole carried *in / down / up* the basket. (Dehé 2002: 4)

(46) *Nicole carried *inside / downwards / upwards* the chairs. (Dehé 2002: 4)

Aspectual particles can sometimes combine with an adverb for intensification purposes (cf. Thim 2012: 17):

(47) It took only a month or so to **fill the jar completely up**. (COCA)

They are sometimes also redundantly used in combination with verbs whose closed aspect is inherent (cf. Thim 2012: 17):

(48) Alex **finished up** his work as he normally did around eleven-thirty. (COCA)

This can sometimes happen with directional particles, too, in pleonastic constructions, although not all of them are acceptable:

(49) I **returned back** here last summer, July 15, got the old band back together. (COCA)

(50) *He **entered** the room **in**. (Thim 2012: 17)

In compositional combinations with a directional particle, fronting of the particle is possible (cf. Thim 2012: 24):

(51) **Up went the kite**, the lantern bobbing beneath it. (COCA)

If the subject is a full noun phrase, fronting takes place with the verb preceding the subject, as seen in (51) above, but if the subject is a pronoun, it will be placed immediately before the verb, as in (52) below:

(52) The door buzzes, Miss Betsy puts her shoulder into it, and **up she goes**, three flights of stairs. (COCA)

Both Dehé (2002) and Thim (2012) note that probably the most fruitful particle used in the aspectual verb-particle constructions is *up*:

> Quite clearly it is *up* that is the most central aspectualizer among the particles, both in terms of frequency and of meaning. (Thim 2012: 18)

Other telic aspectualizers[5] mentioned by Thim are *down, out, over,* and *through*, some of them, such as *down*, for instance, floating on the fuzzy border between directional and aspectual.

Opposed to telic aspectualizers are particles such as *on* and *along*, which are atelic, indicating an ongoing situation (cf. Thim 2012: 18):

(53) He **went on** to say that Obama wants students to attend liberal universities so that they can be inculcated in the same values that drive the president. (COCA)

A particular type of verb-particle construction is what Jackendoff (2002: 80) calls the "**time-*away* construction**". The constituents of this type of constructions are an intransitive verb, followed by a noun phrase indicating a period of time, and the particle *away*. The example below is taken from Jackendoff (2002):

(54) Frank drank the night away.

No other post-verbal noun phrase is allowed in this type of construction, except the temporal phrase, as the example below shows (cf. Jackendoff 2002: 81):

(55) *Frank drank scotch the night away.

Verbs which require a direct object are not allowed in this type of construction:

(56) *Frank devoured the night away. (*Frank devoured.) (Jackendoff 2002: 81)

[5] *Aspectualizer* is an umbrella term used by Thim (2012: 16) for both aspect and lexical marking.

Jackendoff also points out that no other particle than *away* is allowed in this construction. The two examples below cannot be likened on semantic grounds, that is, while in (58) *up* has an aspectual meaning, suggesting the completion of the action, (57) is not acceptable.

(57) *Frank drank the night up.

(58) Frank drank the scotch up.

Jackendoff (2002) notes the appearance, in the last thirty years, of a distinct subclass of idiomatic verb-particle combinations, with the following characteristics:

- The "verb" need not be an actual verb or even an independent word.
- The particle used in this type of constructions is always *out*.
- The core meaning of these constructions is roughly "go into an unusual mental state" (cf. Jackendoff 2002: 73).

Here are a few examples of such recent constructions taken from Jackendoff (2002: 74):

- Intransitives: *flake out, zonk out, zone out, bliss out, flip out, space out, phase out, crump out, veg out, chill out*.
- Transitives: *gross* NP *out, weird* NP *out, creep* NP *out*.
- Intransitive or transitive: *freak* (NP) *out, wig* (NP) *out, stress* (NP) *out, mellow* (NP) *out.*

Jackendoff notes another seemingly recent class of constructions involving *out*, but which does not necessarily involves a verb. To illustrate his claim, he gives the example of a person (himself) who has programmed for six hours in a row. In such a situation, he might say: *I'm (all) programmed out,* or *I('ve) programmed myself out* (cf. Jackendoff 2002: 85). This is basically similar to other already established forms, such as *burnt out*, which seem to have served as a model for this type of construction. A different, fixed idiomatic particle combination is illustrated in the following example:

(59) Harold sang his heart out. (Jackendoff 2002: 86)

Jackendoff notes that the word order cannot be changed in this construction:

(60) *Harold sang out his heart. (Jackendoff 2002: 86)

And manner adverbs cannot intervene between the verb and the noun phrase (cf. Jackendoff 2002: 86):

(61) *Harold sang happily his heart out.

Elenbaas (2007: 25) notes that the verbs in verb-particle constructions are chiefly monosyllabic light verbs, which allow multiple particle combinations, as shown in Table 4.

Table 4. Combinational variants of monosylabic light verbs with various particles (based on Elenbaas 2007)

Light verb	Particle variants
come	*up / in / down / out / off*
get	*up / in / down / out / off*
give	*up / in / out*
go	*up / in / down / out / off*
put	*up / down / out / off*
take	*up / in / down / out / off*

However, verb-particle constructions are not restricted to this type of verb. Any verb which expresses the means or manner by which the action is performed can enter a verb-particle combination (cf. Elenbaas 2007: 25). Thus, 'means' verbs indicate the tool with which the action is performed, e.g. *belt up, chalk up, cork up, dish up, lace up, mop up, polish up, elbow out, hammer out, iron out, nose out, clock out, fence off, mouth off*, while manner verbs indicate the way in which the action is carried out, e.g. *bundle up, carve up, shake up, sharpen up, freeze out, portion out, rub out, chop down, wipe off, etc.* (cf. Elenbaas 2007: 25). Sometimes, the particle itself is used as a verb, as in the examples below, taken from Elenbaas (2007: 26):

(62) Mum offed the television.

(63) The butcher upped the prices.

(64) The hooligan downed the beer.

Far from being exhaustive, this presentation has given an overview of the most important syntactic features of the verb-particle combinations in English. The remainder of this chapter

is exclusively dedicated to a discussion on the phrasal verbs with *up* and *out*, and their use in the two main varieties of English, as evidenced by corpus findings.

5.4. *Up*

As I mentioned in the introductory part of this chapter, the main focus in this section will be on those phrasal verbs constructed with the two most productive particles, i.e. *up* and *out*. Accordingly, in the two next sub-sections I will first take a closer look at the various semantic values the particle *up* can have, and second, I will review my own empirical data related to the verb-particle constructions with *up* retrieved from the corpora.

5.4.1. Values of *Up* in the Literature

Like Dehé (2002) and Thim (2012), Kennedy (1967), too, argues that *up* is by far the most common particle used in verb-particle constructions, occurring "in more than twice as many combinations as the particle *out*, the next in frequency of combination" (Kennedy 1967: 23). These claims are corroborated by *Collins COBUILD Dictionary of Phrasal Verbs* (1989: 449), which lists *up, out, off, in, on,* and *down* as the most common verbal particles, exactly in this descending order of frequency, the combinations with *up* and *out* constituting 28% of the phrasal verbs listed in the dictionary. At the opposite extreme, according to the dictionary, particles such as *aback* and *across* seldom occur in verb-particle combinations. Moreover, *up* seems to have been the most widely spread particle since the very dawn of the English language (cf. Hiltunen 1983: 208). Hiltunen bases his claim on the findings in his corpus compiled from some of the most important Old and early Middle English prose texts. He maintains that in Old English, *up*, also spelt *upp* at that time, had mostly an adverbial value, but he could however find two instances of prepositional use, too. He notes that the prepositional use of *up* became more common in the Middle English period. For a more comprehensive picture and a better understanding of the current status of *up*, I will start with some information related to its use in Old English, provided by Hiltunen (1983), as this offers a valuable, in-depth insight into the beginnings and subsequent development of *up*.

5.4.1.1. Historical Values

A first combination of *up* in Old English, mentioned by Hiltunen, is with the verb *aræran*, together signifying 'bringing up', 'raising up' (cf. Hiltunen 1983: 208). An example of this

combination can be seen in the example below, found by Hiltunen (1983: 123) in the late Old English text *Apollonius of Tyre*:

(65) ac *arærde* hi *up.*

‘but brought/raised them up’ (*Apollonius of Tyre*, late OE).

A slight extension of this early meaning of *up* can be found in an early Middle English text (Hiltunen 1983: 208):

(66) if ani cumþ and bri[n]gþ tidinges of idelne(s)ses, and is spekende sotwordes ðe *aræreð up* hleitres …

‘if anyone comes and brings tidings of idleness, and is speaking words which bring up laughter’ (*Vices and Virtues*, c. 1200 AD).

Another verb with which *up* combined in Old English is *blawan*, according to corpus evidence found in the collection of sermons *Ælfric’s Lives of Saints* (cf. Hiltunen 1983: 208). The modern meaning of this construction is ‘blow up’, ‘explode’. An extension from the original, literal meaning is found in the combination *bringan up*, as shown by the following two examples (cf. Hiltunen 1983: 209):

(67) and hwon his blowinge ne geineð nout: þeonne *bringeð* he *up* sum luðer word.

‘and when his blowing is of no avail, he then brings up some insulting word’ (*Ancrene Riwle/Wisse*, early 13th century AD);

(68) or þat may noght be halden withouten syn; to lede karols; to *bryng up* new gyse; to be rebell agayne hys soverayns

‘or that may not be held to be without sin; to lead carols; to introduce a new fashion; to be a rebel against his sovereign’ (*R. Rolle*, 1300-1349).

In (67), the meaning is ‘suggest’, ‘bring (into conversation)’, while in (68), *bringan up* is used with the meaning ‘introduce (a new fashion)’ (cf. Hiltunen 1983: 209). Hiltunen also points out that unlike *bringan up*, the meanings of *beran up* and *bregdan up* remained literal in early Middle English. A frequent construction, with the meaning ‘land’, found in Old English, particularly in *The Anglo-Saxon Chronicle*, is *cuman up*, as seen in (69) below (cf. Hiltunen 1983: 209):

(69) þa *common up* on Limene muþan. mid ccl. hunde scipa.

‘then came up into the Lime estuary with 250 ships’ (*Anglo-Saxon Chronicle*).

Hiltunen holds that this combination may have acquired a more abstract meaning during the early Middle English period, namely that of ‘come about’, ‘arise’, as seen in the two examples below (cf. Hiltunen 1983: 209):

(70) amidde ðe redunge hwon ðe heorte likeð wel, Þeonne *cumeð upp* adeuociun

‘in reading, when the heart feels delight, devotion arises’ (*Ancrene Riwle/Wisse*);

(71) affectiun is hwon ðe þouht geð inward & ðe delit *kumeð up*

‘affection is when the thought goes inward and fondness comes up/is excited’ (*ibid*.).

Hiltunen maintains that the origin of the meaning in (70) and (71) may lie in the meaning of *cuman up* in (72), namely ‘spring up (out of the ground)’ (cf. Hiltunen 1983: 209):

(72) þa ne *com* ðær nænig grownes *up* ne wæstm

‘then there came up no growth nor fruit’ (*The Anglo-Saxon Chronicle*).

However, Hiltunen finds some instances of the combination, with a very similar meaning, in Old English, too (cf. Hiltunen 1983: 209):

(73) ‘Wa us, forþon þe þas ealle *up coman* for þissum ælþeodigum

‘Woe betide us, because these all came up for/due to these foreigners’ (*Blickling Homilies*; late 10th century AD).

While combinations such as *faran up*, and *gan up* maintained their literal meanings, *don up* sometimes signified ‘exhume’ in both Old and early Middle English, as in the following example retrieved from the early Old English text *Bede’s Ecclesiastical History of the English People* (cf. Hiltunen 1983: 209):

(74) þæt heo woldan his ban geneoman 7 *up* ofer eorðan *adon*

‘that they wanted to dig up his bones and remove them above ground’ (*op.cit.*, 9th century AD).

The construction *giefan up*, meaning ‘give up’, became a lexicalized item, according to Hiltunen’s research, around 1150 A.D., and a similar meaning can also be found in *ȝeldan up*, as illustrated in (75) below (cf. Hiltunen 1983: 210):

(75) and þreateð ðet ȝe wulleð *ȝelden up* þene castel

‘and threaten that you would yield up the castle’ (*Ancrene Riwle/Wisse*).

Hiltunen notes that the meaning of ‘castel’ here is not literal, as one might be tempted to believe, but metaphorical, standing for ‘a spiritual castle’, i.e. ‘religion’.

According to Hiltunen, *lætan up* is an instance of a construction in which a non-finite verb may have been omitted, and at the same time, a good example of how phrasal verbs are sometimes created. For example, in (76), the meaning of *leten up* is specifically ‘let somebody climb on one’s back’ (cf. Hiltunen 1983: 211):

(76) buh ðe he seið adun: and *let* me *up*

‘“bow down”, he says, “and let me [climb] up”’ (*Ancrene Riwle/Wisse*).

Leten in this case undertook the meaning of the omitted non-finite verb, i.e. *climb.*

Locian, logian, niman, and *macian* are verbs with which *up* has its original, directional meaning (cf. Hiltunen 1983: 211). However, in (77) below, *niman up* is synonymous with *don up* mentioned a little earlier, that is, its meaning is ‘exhume’ (cf. Hiltunen 1983: 211):

(77) þ hi *moston nyman up* of þam byrgene þone arceb sc̃e Ælfheah

‘that they had to take up from the grave the archbishop, St. Ælfheah’ (*Anglo-Saxon Chronicle*).

In (78), *settan up* has, according to Hiltunen (1983: 211), the meaning ‘set up as a token’:

(78) Hwæt ða Moyses for ðam folce gebæd, and God þærrihte bebead Moyse þæt he geworhte ane ærene næddran, and *sette up* to tacne

‘Then Moses prayed for the people, and God commanded Moses to have a serpent made of copper, to put/set up as a sign’ (*Bible*: Numbers 21).

Hiltunen notes that *Oxford English Dictionary* quotes this excerpt from *Ælfric’s Catholic Homilies* as the earliest instance of the phrasal verb *set up* with the meaning ‘erect’. But Hiltunen found an even earlier combination of *up* with *asettan,* having the meaning ‘move one’s self’, in *The Anglo-Saxon Chronicle* (cf. Hiltunen 1983: 211):

(79) 7 wel gehwær hergedon 7 bær[n]don, swa þ hy *upp asetton* on ænne siþ

‘and harried and burned everywhere, so that they set out on a journey’ (*Anglo-Saxon Chronicle*).

Two other verbs with which *up* was commonly combined in Old English are *(a)springan* and *weorpan*, as illustrated in the following two examples (cf. Hiltunen 1983: 211):

(80) þat non godes word *upp* ne mai *springen*

’that no word of God may spring up’ (*Vices and Virtues*);

(81) Æfter þæm þe hie adruncne wæron, hie *wearp* se sæ *up.*

‘after they were drowned, they were thrown ashore by the sea’ (literally, ‘the sea threw them up’; the Old English *Orosius*, late 9th or early 10th century AD).

Hiltunen (1983: 212) notes that Old English *up* was also part of the prepositional items *uppan, onuppan*, and a constituent of the adverbial *uppe.* The Middle English *upon*, which gradually became a common variant of *on* in that period, has its roots in both *uppan (uppon)*, and the collocation *up (…) on*, in which an element was sometimes interposed between *up* and *on* (cf. Hiltunen 1983: 212).

5.4.1.2. Contemporary Values

Coming back to the present, I will further discuss the most common contemporary *up*-combinations, as found in some of the most notable studies in the literature. As already mentioned in this chapter, the default meaning of *up* is literal and has to do with a **movement upwards**. A few good examples of this basic use of *up* are given by Kennedy (1967: 23): *brick up* (a well), *build up, pick up, pile up, prick up* (of an animal’s ears), *raise up, rise up, rouse up, run up* (a flag), *spring up, stand up, take up* (literally, e.g. a suitcase). Kennedy points out that, while in some of the above examples the addition of *up* is somewhat redundant since the meaning of the verb alone already has by itself to do with the upward motion, e.g. *raise, rise*, in some others, the particle is compulsory, e.g. *spring up* (of a plant), or *stand up* (of a person).

Another major class of uses of *up* mentioned by Kennedy is what linguists now call idiomatic or non-compositional verb-particle constructions, where the meaning is no longer literal, in such combinations as *add up, balance up, bid up, blow up* (with the meaning ‘to boast’), *call up, check up, cheer up, choose up, count up, fill up, fire up, grow up, hang up, keep up, live*

up, mop up, offer up, paste up, praise up, ring up, set up, shake up, sign up, stir up, sum up, wake up (cf. Kennedy 1967: 23). In combinations such as *back up, hit up, hurry up, move up, pull up* (with the meaning 'to move ahead'), *speed up, start up, steam up* the implied movement is not upwards, but rather **forward** (cf. Kennedy 1967: 23). The meaning inferred from such combinations as *bring up* (with the meaning 'to cause (something) to be of a required standard'), *catch up, connect up* (cables, pipes, etc.), *join up, lead up, line up, link up, match up, measure up, meet up, size up* (with the meaning 'to conform to certain specifications of dimension') is a **movement to a certain position or standard** (cf. Kennedy 1967: 24). Combinations such as *act up, divide up, give up, hunt up, look up, open up, serve up, show up, speak up, study up, think up, turn up, write up* suggest the idea of **bringing into prominence or consideration** (cf. Kennedy 1967: 24). Another class of combinations proposed by Kennedy is that of verbs which suggest the idea of **making tight or secure**, e.g. *bind up, board up, bottle up, box up, bundle up, button up, cage up, close up, cover up, fasten up, hook up* (a dress), *lace up* (a shoe), *lock up, nail up* (a box), *patch up, plug up, rope up, screw up, seal up, sew up, shut up, solder up, strap up, tie up, tighten up, wall up, wrap up* (cf. Kennedy 1967: 24). A group of combinations somewhat related to the previous one is that of verbs which suggest the idea of **bringing together**, e.g. *bunch up, collect up, fold up, gather up, make up* (a collection), *pack up, roll up, stock up, store up* (cf. Kennedy 1967: 24). The second commonest value of *up*, after its literal meaning of elevation or movement upwards, is the perfective one, suggesting the completion of the action expressed by the verb or, as Kennedy puts it, suggesting "the idea of '**bringing to or out of a certain condition**'" (Kennedy 1967: 24). The number of such possible combinations is virtually unlimited, as aspectual *up* can combine with almost any verb. However, among the most common combinations, Kennedy mentions verbs such as *beat up, black up, blister up, brighten up, bruise up, brush up, build up, carve up, clean up, clear up, clog up, cloud up, cook up, curl up, do up* (with the meaning 'to exhaust a person'), *dress up, dust up, fatten up, fit up, fix up, flush up, freeze up, freshen up, grease up, heat up, light up, loosen up, mark up, mess up, mix up, peel up, polish up, powder up, read up, rest up, rub up, rust up, scratch up, shape up, shine up, skin up, slash up, smell up, smoke up, smooth up, splash up, spot up, stick up, tangle up, thicken up, tidy up, tighten up, touch up, trim up, twist up, warm up, wash up, wet up* (cf. Kennedy 1967: 24).

In addition to the idea of completion suggested by Kennedy, Bolinger (1971) associates aspectual *up* with a "notion of **closing a gap** between the eye of the viewer and the thing

viewed" (Bolinger 1971: 98). The following examples given by Bolinger illustrate this nuance of aspectual *up* (*ibid.*):

(82) He came up (to me) and said …

(83) She walked up (to the door) and knocked.

(84) Pull up a chair and join the game.

Bolinger notes that the use of *down* instead of *up* in (82) would imply an actual downward motion in the meaning, but with *up* the meaning is perfective, more precisely a goal oriented closing in. Another nuance of the perfective meaning of *up* is that of **resultant condition**, as illustrated in the examples below (cf. Bolinger 1971: 99):

(85) The ice broke up.

(86) She scratched up his face.

Not only completion can be expressed by perfective *up*, but sometimes also **inception**, as in *She took up dancing* (cf. Bolinger 1971: 99). In (87) and (88) below the inferred meaning is that of "attaining a high **intensity**" (Bolinger 1971: 99):

(87) Let's brighten up the colors.

(88) Speed up the engine.

Bolinger points out that it is difficult, if not impossible, to draw a clear-cut line between all these meanings. *Grow up*, for instance, is perfective, but also directional, while *give up*, although perfective, still keeps something of its original, directional meaning, which makes combinations of synonymous verbs such as *to abandon* with *up* ungrammatical (cf. Bolinger 1971: 101). This is illustrated in the two examples below taken from Bolinger:

(89) He gave up the property and goods.

(90) *He abandoned up the property and goods. (Bolinger 1971: 101)

In *she made up an excuse*, the meaning is closely related to that of inception, namely 'to improvise something', or 'to cause something to appear' (cf. Bolinger 1971: 101).

In a cognitive linguistic approach, Tyler & Evans (2007) term the primary meaning of a particle a 'proto-scene':

> A proto-scene is an idealized mental representation across the recurring spatial scenes associated with a particular spatial particle; hence it is an abstraction across many similar spatial scenes. (Tyler & Evans 2007: 52)

According to Tyler & Evans, humans segment their perceptions of the world into spatial scenes. In a phrase such as *the cup is on the table*, for instance, the spatial scene is conceptually constructed and understood as a cup and a table sharing a particular spatial relation in which the cup is in direct contact with the table, and the table functionally supports the cup (cf. Tyler & Evans 2007: 27). In their theory, Tyler & Evans describe a spatial relation as being made up of a trajector (TR) and a landmark (LM), where the trajector is surrounded by the landmark (cf. Tyler & Evans 2007: 25). In the above example, for instance, the trajector is the cup, and the landmark is the table.

In the case of *up*, the proto-scene "denotes a relation in which the TR is directed towards the top of an oriented LM" (Tyler & Evans 2007: 136). It is implied that the landmark has a top and a bottom and that the trajector is oriented and undergoes a vertical elevation, from the bottom towards the top. Tyler & Evans hold that a cluster of three other senses, which have nothing inherently spatial, and together are termed the Quantity Cluster, derives from the proto-scene of *up* (cf. Tyler & Evans 2007: 138). The first of these additional meanings is what Tyler & Evans call '**the more sense**' (Tyler & Evans 2007: 138):

(91) The maid plumped up the cushions.

(92) Pump/turn up the volume/heat.

In the two examples above, the notion of an increase in vertical elevation is correlated with a notion of an increase in quantity. The result of plumping up a cushion is an increase in its volume. Similarly, pumping up the volume or turning up the heat results in a higher volume level or more heat.

The second sense of *up* within the Quantity Cluster is '**the improvement sense**' (cf. Tyler & Evans 2007: 139). Tyler & Evans explain that an increase in quantity or more of something

can sometimes be associated with enhancement, betterment, improvement. Earning more money, for instance, implicates a better standard of living.

(93) Dave and Kirsten decided to get dressed up and go to a nice restaurant. (Tyler & Evans 2007: 139)

Tyler & Evans explain that the interpretation of (93) above is that the two characters decided to put on stylish and not casual clothing, with the goal of improving their appearance. This distinction in meaning from the choice of using the simple verb *to dress* is made by the particle *up*. However, Tyler & Evans note that this nuance of improvement added by *up* does not imply putting on more clothes, but more elegant ones.

The third sense of *up* within the framework of the Quantity Cluster, is one that has been discussed before, namely '**the completion sense**' (cf. Tyler & Evans 2007: 139), but Tyler & Evans go into even more detail in the description of this meaning of perfective *up*.

(94) The waiter filled up the mug with beer.

(95) Be sure to gas up the car for the trip.

(96) Let's load up the truck and get going. (Tyler & Evans 2007: 140)

In (94) above, the implication is that the beer in the mug has reached the latter's full capacity, and therefore, the action of filling it is completed. In (95) the situation is different, in the sense that although we are still dealing with the same action of filling a container, i.e. a car's fuel tank, up to its full capacity, the process has not been completed yet. Furthermore, the interpretation of (96) is different from both (94) and (95), in that the action will go on until the task of loading the truck is complete. Tyler & Evans point out that the completion sense associated with *up* is conventionalized, and this can be proved by the following two examples:

(97) Let's finish up this work today.

(98) They closed up the shop for the night. (Tyler & Evans 2007: 140)

Tyler & Evans point out that there is no clue which can be inferred from the context of the two examples above as to how *up* is compatible and combined with verbs such as *finish*, or

close. The only possible explanation is that *up* has a conventional completion sense associated with it.

A curious nuance of the completion sense, which contradicts the one discussed so far, is illustrated in the examples below:

(99) The flashing won't work. We must have used up the batteries.

(100) Students, turn in your papers; your time is up!

(101) The guests drank up the wine and promptly fell asleep. (Tyler & Evans 2007: 140)

Tyler & Evans point out that the completion sense in these examples implies a notion of depletion rather than filling a container to capacity. Thus, the interpretation of (99) is that all the power stored in the batteries has been consumed. In (100), the implication is that the time allocated to the students to write their papers has elapsed, and that the students have used all the allotted time. In (101), the guests have consumed the whole amount of available wine. The notion of depletion denoted by these instances of *up* seems contradictory to the general meaning of *up* indicating completion through filling a container up to its full capacity. However, Tyler & Evans point out that the consequence of depletion eventually comes to be one of completion. For instance, in (101), the consumption of the whole amount of wine will cause the activity of drinking to be completed. Tyler & Evans conclude that there are actually two distinct completion senses of *up*: one resulted from filling a container up to its full capacity, and hence the filling activity being complete, and the other, resulted, on the contrary, from the depletion of a container or some other entity (cf. Tyler & Evans 2007: 141). In both cases the activity is complete.

Elenbaas (2007) proposes five meanings of *up*: a core directional meaning of **upward motion** as in *throw up* (a ball), and four non-transparent meanings derived from the primary one, namely **emergence,** as in *show up, wake up*, **initiation,** as in *start up* (a car), **cessation**, e.g. *slow up*, and **completion,** as in *clean up, warm up* (cf. Elenbaas 2007: 21). Elenbaas notes that sometimes, the meaning of the verb and the particle overlap, as in *start up*, for instance, where both the verb and the particle indicate initiation. On the other hand, two different, opposing particles may sometimes indicate the same meaning, as, for example, in *slow up* and *slow down*¸ where both *up* and *down* indicate cessation.

Collins COBUILD Dictionary of Phrasal Verbs (1989: 487) lists 12 meanings of *up* grouped as follows: the literal meaning of **movement and position**, e.g. *jump up, run up* (a hill). Combinations which mean 'lift' are also included here, in which case *up* indicates the movement upwards, while the verb expresses the manner in which the movement is performed, *e*.g. *pick up, dig up*. The combinations which indicate an increase in quantity or intensity are grouped under the meaning **increase and intensification**, e.g. *brighten up, build up* (ice on the inner walls of a deep freezer), *come up* (wind, applause), *fatten up, go up* (prices, expenditure, etc.), *grow up, heat up, hurry up, level up, light up, mark up, mount up, move up, open up* (with the meaning 'accelerate'), *pick up* (trade, business, economy), *pile up, puff up, push up, sharpen up, shoot up, sing up, speed up, speak up* (i.e. 'speak louder'), *step up, stir up, swell up, turn up* (radio volume control, heat, etc.), *warm up, whip up* (sympathy, hatred, etc.), *work up* (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 487).

The **improvement and preparation** group reunites combinations such as *boil up, dress up, fit up, fix up, make up*, e.g. *make up the bed, set up, touch up* (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 487). *Boil up water*, for instance, implies that the water is being boiled in order to be ready to use, hence the preparation meaning. As for the improvement meaning, the instance of *dress up* has already been discussed. Combinations such as *bandage up, belt up, bind up, block up, bottle up, brick up, button up, chain up, clog up, cork up, earth up, fasten up, freeze up, hook up, join up* (paper clips in a chain), *lace up, lock up, nail up, pin up, seal up, sew up, stitch up, stop up, strap up, string up, tangle up, tie up, tighten up, wall up, zip up* are included in the **fastening and restriction** group of meanings, in which the combinations imply that some entity is being fastened or restricted in a certain manner (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 488). Under the **approach** meaning, the dictionary lists those combinations which imply that two entities are moving closer, or are staying close together, e.g. *catch up, catch up with, creep up on, keep up, keep up with, line up, match up, square up* (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 488).

Combinations which suggest the idea of something being spoilt or damaged are listed under **disruption and damage**, e.g. *blow up, botch up* (a job), *crack up, foul up, mess up, play up* (of a machine), *smash up* (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 488). Some of these combinations, such as *botch up*, for example, refer to a person who performs a piece of work in an unsatisfactory manner. **Completion and finishing** reunites such

combinations as *add up, beat up, burn up, clean up, cover up, drink up, dry up, eat up, end up, fill up, finish up, hang up, mash up, mop up, pump up, settle up, size up, sum up, swallow up, use up, wrap up, write up* (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 488). As already discussed, some of these combinations suggest the completion of the action, while some others imply a notion of depletion or consumption, also having completion as a consequence. **Rejection and surrender** are suggested by *up*-combinations such as *cough up* ('pay'), *deliver up, give up, pass up* (an offer), *pay up, sell up* (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 489). Phrasal verbs such as *brew up* (of a storm), *bring up, crop up, draw up, dream up, spring up, think up* imply **happening or creation** (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 489). For instance, if something is cropping up, it means that it is happening, it occurs, often unexpectedly, while drawing up a plan implies an act of creation. Under **collection and togetherness** are listed those combinations of *up* which imply that several objects are being gathered together, and that sometimes, the objects are compacted or shrunk in the process. Quite often, the verb denotes the type of container in which the things are collected, e.g. *bag up* ('put things in a bag'), *bunch up, bundle up, collect up, crate up* (bottles in a crate), *gather up, join up, link up, pack up, pair up, queue up, ring up, round up, sign up* (of people; with a website, an organization, etc.; in this case, a notion of involvement is implied), *stock up, store up, sweep up, team up, tie up, wire up, wrap up* (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 489). The combinations which convey the idea that some piece of information has been revealed or discovered, are listed under the **disclosure** meaning, e.g. *dig up* (a secret; 'the secret has been discovered'), *drag up, hunt up, own up* ('confess, reveal'), *show up, throw up, yield up* (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 490).

Finally, *Collins COBUILD Dictionary of Phrasal Verbs* lists those phrasal verbs with *up* which imply that an entity is moving away, or is being separated from another, under the meaning **separation**, e.g. *break up, chop up, cut up, divide up, parcel up, slice up, split up* (cf. *Collins COBUILD Dictionary of Phrasal Verbs* 1989: 490).

Before moving forward to the practical part of this section, I will try to summarize the intricate semantic interplay of *up* with the multitude of verbs with which it combines. Despite the various terminologies adopted by various linguists to describe its values, the meanings of *up* are roughly the same in all theories, irrespective of any particular classification. Thus, the basic meaning is literal, namely it has to do with upward motion. All the other, non-

transparent senses are derived from this primary meaning, called the proto-scene of *up* by Tyler & Evans. Historically, the literal meaning dominated the whole Old English period. Slight abstract extensions of the primary meaning started to be used in the Middle English period, as evidenced by Hiltunen in his corpus-based research. A peculiarity of the present-day literal meaning of *up* observed by Kennedy is that the implied movement is not always upwards, but sometimes forward, as in *hurry up* or *speed up.* In combinations having transparent meaning, with transitive verbs, that is, combinations which mean 'to lift something', *up* signals the upward motion, while the verb indicates the manner in which the lifting is carried out. When not used in its literal sense, *up* has a perfective function, indicating that the action expressed by the accompanying verb is complete. We have seen that in certain combinations, completion can be achieved in two different, seemingly contradictory, ways: by filling some container up to its full capacity, or on the contrary, by its depletion or consumption, which eventually leads to the same completion of the action. Together with the completion sense, a notion of increase in quantity and one of improvement which can be inferred from certain *up*-combinations, make what Tyler & Evans term the Quantity Cluster of *up*. The approach meaning listed in *Collins COBUILD Dictionary of Phrasal Verbs* suggests a gap closing between two entities, which Bolinger also refers to. Some combinations suggest the idea of restriction or fastening, while another distinct set of constructions indicate things being collected or brought together. Opposed to the latter category are those combinations which indicate that things are separated from one another. In some combinations involving perfective *up,* the emphasis is on the result of the action, while in others, the suggested meaning is that of initiation of the action, and not its completeness. Closely related to the quantity increase meaning is a sense of *up* which suggests that a high intensity has been reached. We have seen that the completion meaning of *up* is conventionalized, and this is the reason why some otherwise unlikely combinations occur. Last but not least, like other particles, *up* can be used alone as a verb in its own rights, but of course, in this case, we can no longer talk about a phrasal verb.

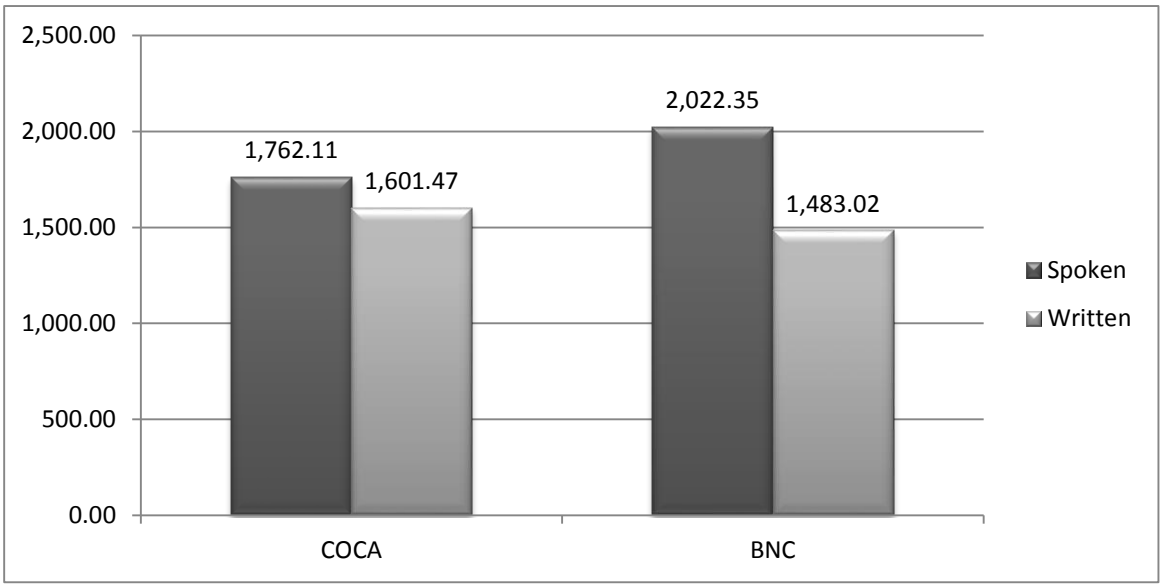
5.4.2. Corpus Findings

5.4.2.1. Continuous VPCs with *up*

Continuous VPCs are those verb-particle combinations in which the particle is adjacent to the verb and the direct object follows the particle. It is this type of VPCs with *up* that I will focus on in this subsection.

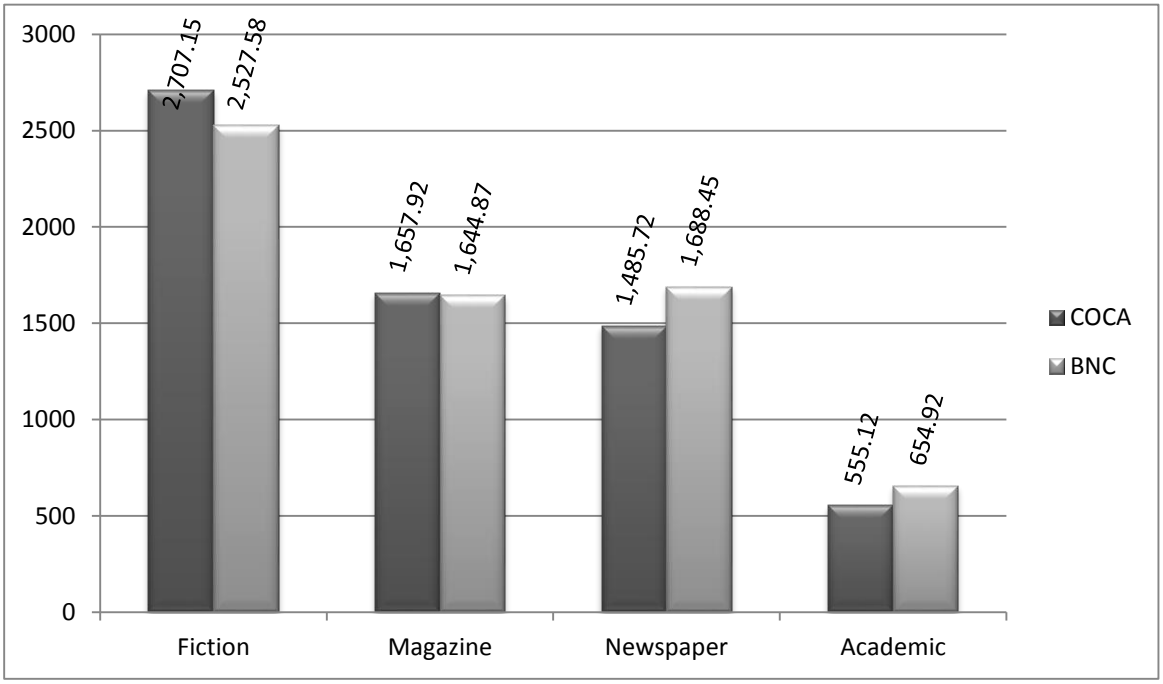
In order to prevent the auxiliaries *be*, *do* and *have* from appearing in the results, which in some cases would lead to completely irrelevant matches that have nothing to do with phrasal verbs, I used the following query syntax: [vv*] *up*. This tag set returns only those matches which include a lexical verb. A first striking detail which can be observed with respect to the use of the continuous VPCs with *up* in British and, particularly in American English, is the minor difference between speech and writing. If in the case of contractions and colloquial *like*, the frequencies in speech are much higher than in writing, in the case of the continuous VPCs with *up* the difference is hardly noticeable, especially in American English, as Figure 56 shows. Curiously enough, the continuous VPCs with *up* seem to be more widely used in British English than in American English, in spoken language. On the other hand, the overall high frequencies of this type of VPCs are not surprising, given the general popularity of phrasal verbs.

Figure 56. Frequencies per million words of the continuous VPCs with *up* in COCA and BNC, in the spoken vs. written material



In terms of genre, the continuous constructions seem to be preferred, according to the corpus evidence, by writers of fiction, in both British and American English, with outstanding frequencies per million words of 2,527.58 and 2,707.15, respectively, and with a minor difference between the two varieties (Figure 57). The magazine and newspaper genres have roughly similar frequencies in both varieties of English, while the academic genre occupies the last position in the chart, also with a minor difference between British and American English, in favor of the former. The poorer representation of the continuous VPCs with *up* in the academic genre could be explained by the general association of phrasal verbs with informality, in the context of the well-known conservatism and highly formal style of academic prose. It is very likely that authors of academic texts opt for the more formal, lexical equivalents of the phrasal verbs when they make their choices in drafting an academic work. However, the frequencies of 555.12 for American English and 654.92 for British English are not insignificant, especially when compared to the frequencies of the highly informal contractions discussed in a previous chapter.

Figure 57. Normalized frequencies of the continuous VPCs with *up* in the written material of COCA and BNC, divided by genre



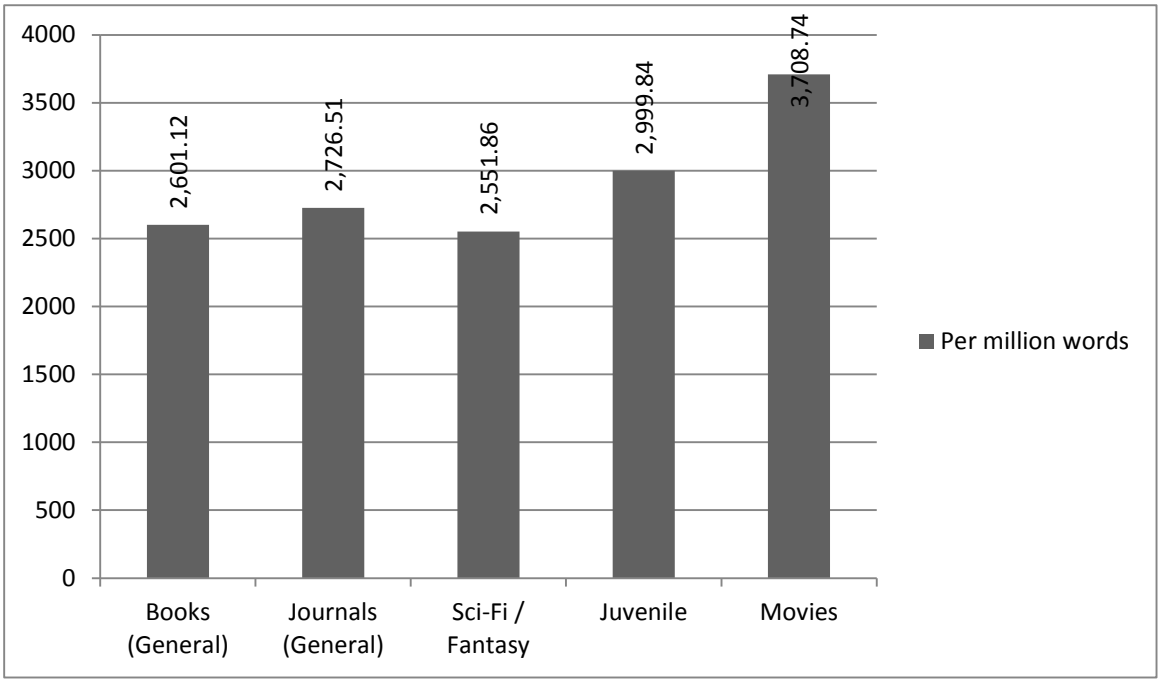
In (102) below, extracted from an academic text found in COCA, the author opted for the use of the phrasal verb *take up* instead of a more formal choice such as *occupy*:

(102) The poster headline read "Adolph Hitler" while his picture **took up** most of the poster. (COCA)

The use of verb-particle constructions in academic prose is likely to be part of a more complex context in which an informal style is deliberately used. This might be particularly true for academic texts targeted at a readership consisting either of students or non-professional readers, e.g. textbooks, applied and popular science texts, etc. The purpose of this deliberately informal style is to make the texts sound "chatty and user-friendly" (Fowler 1991: 63). A further technique used by writers of academic prose to help their texts gain a wider audience and to make them sound more appealing to the target readership is, according to Hundt & Mair (1999: 230), a simulated direct writer-reader interaction, achieved through devices such as question-answer sequences, urges, e.g. *let us* and its contraction *let's*, or imperatives, such as *take a look at.*

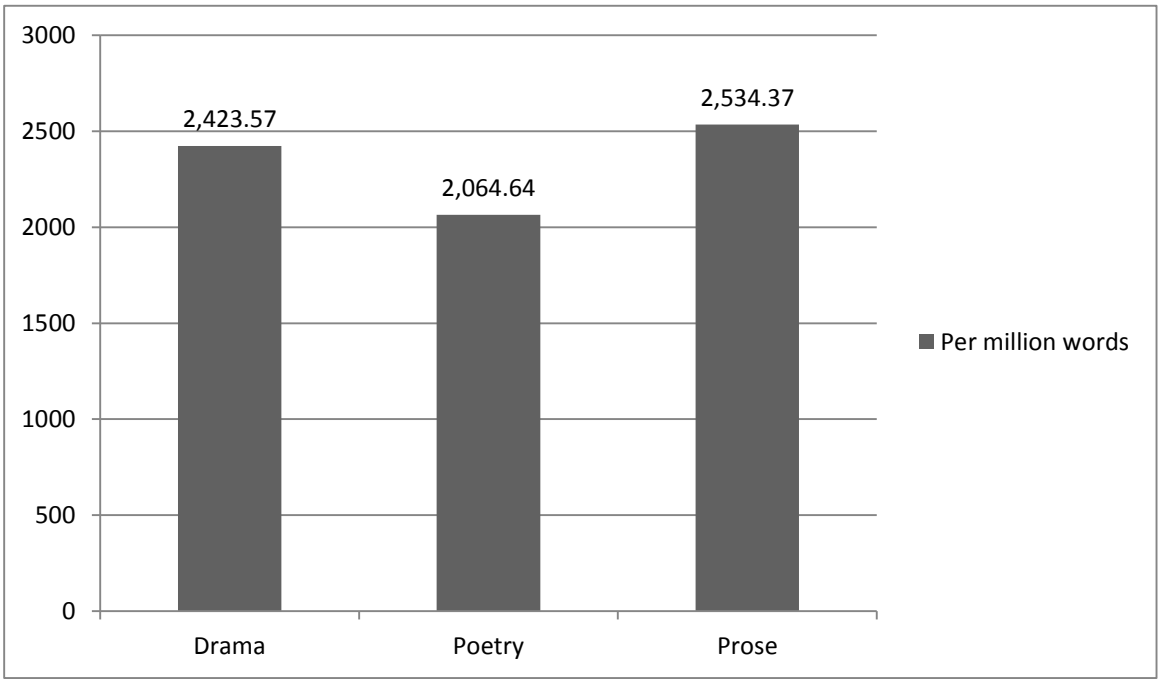
A more detailed look into the fiction section of COCA reveals the distribution of the VPCs with *up* in the subgenres of this category (Figure 58). Thus, the highest frequency is found in movies, with a stunning figure of 3,708.74. The next subgenre in the frequency chart is youth literature, with a normalized frequency of 2,999.84, followed by journals targeted at a general audience, with 2,726.51; general audience books with 2,601.12; and science fiction and fantasy literature, with 2,551.86. The extremely high frequency in movies is fully explicable, since movies are targeted at the popular mass culture. Unfortunately, the structure and categorization of the genres and subgenres in the BNC is different from COCA; hence the incomparability of the two corpora, but even so, a look into the subgenres of the fiction section of the BNC reveals further interesting facts about the distribution of the continuous VPCs with *up* in this genre.

Figure 58. Distribution of the continuous VPCs with *up* in the subgenres of Fiction in COCA



As Figure 59 shows, the BNC provides three subgenres of fiction, namely drama, poetry and prose, of which the last one tops the frequency chart, with a normalized value of 2,534.37, followed by drama with 2,423.57, and poetry, with a frequency per million words of 2,064.64.

Figure 59. Distribution of the continuous VPCs with *up* in the subgenres of Fiction in BNC



What is interesting here is the high frequency of continuous VPCs in poetry, a genre which although not necessarily formal, is however artistic. It is interesting how phrasal verbs are integrated in the poetic language. Here is an example, extracted from the BNC, where phrasal verbs are used in poetry:

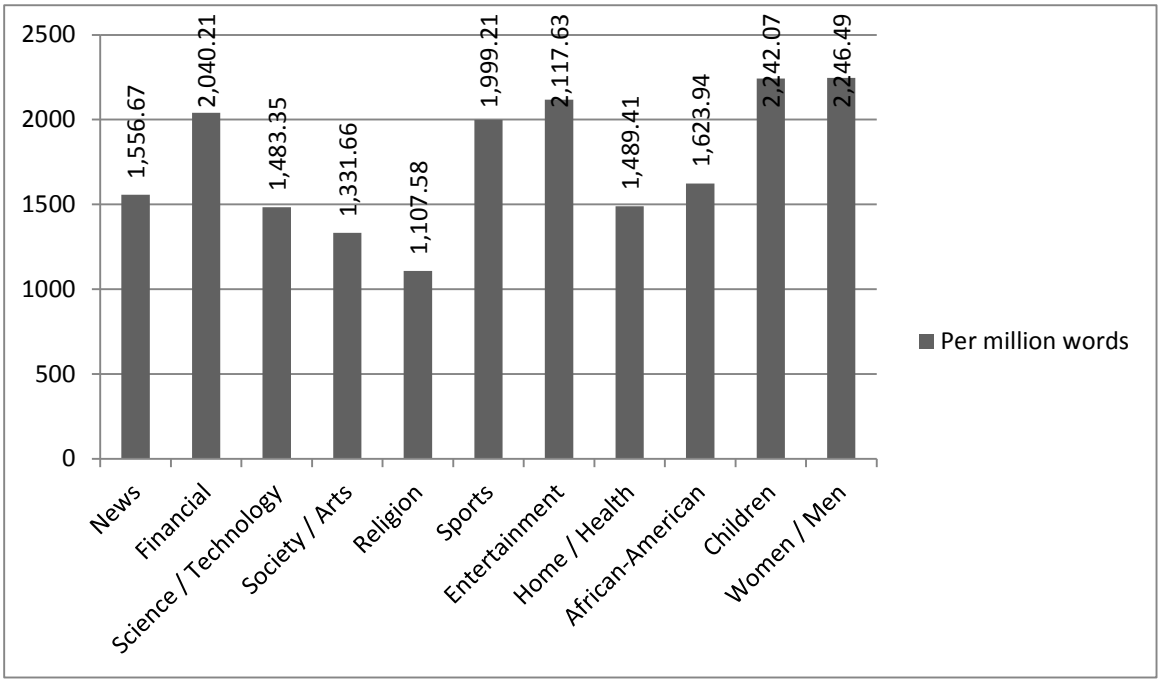
(103) # Another sound **breaks** **up** my dreams, # While crackling flames **eat** **up** the beams # Much louder than a pulley squeaking, # Dirtier than a chimney reeking -- # (BNC: CAV)

Figure 60 shows the distribution of the continuous VPCs with *up* in the subgenres of the magazine category of COCA. Women and men magazines lead the frequency chart, with a normalized frequency of 2,246.49, closely followed by children's magazines, with 2,242.07. This high frequency in children magazines is interesting, as idiomatic constructions are normally acquired at an older age than is basic language, even by native speakers of English. The next positions in the frequency chart of the magazine subgenres are occupied, in order, by entertainment, financial, sports, magazines targeted at an audience made up of African American ethnics, news magazines, home and health magazines, science and technology, society and arts, and finally, religion. The high frequencies in sports, entertainment and finance are not surprising, as these genres are inclined to a more informal style. Nor are the lower frequencies in the news, science and technology, arts and society, and religious magazines surprising, since these genres are, by their very nature, prone to a more formal style. What is surprising here however is the relatively low frequency in the African American ethnic magazines compared, for instance, to the high frequency of continuous VPCs in children's magazines. African American Vernacular English is known for its extensive use of idioms and slang, and this variety of English is even a lender of such constructions to other American English dialects, such as Southern American English, and even General American. Below is an example of a VPC employed in a religious article on the evil of terrorism, extracted from COCA:

(104) Terrorist movements have **cropped** **up** not only in Syria, Iraq and Libya, but in Germany, Italy, Korea and Japan. (COCA)

Here, the idiomatic continuous VPC *crop up* has been used as an alternative to the more formal alternatives *occur, appear* or *emerge*.

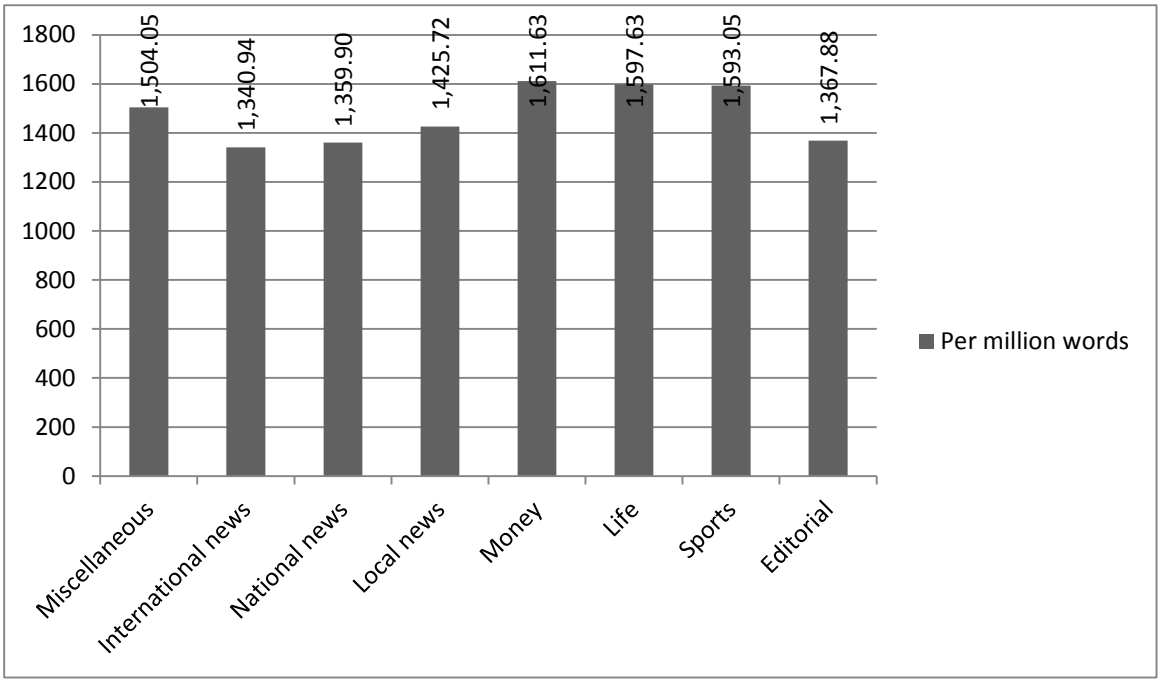
Figure 60. Distribution of the continuous VPCs with *up* in the subcategories of Magazine in COCA



Unfortunately, the BNC does not provide a further classification of its magazine section, which would have made possible a comparison of its subgenres to the magazine subcategories in COCA.

As I mentioned in the introductory part of this chapter, Hundt and Mair (1999) found an increased use of phrasal verbs, especially of the ones constructed with the particle *up*, in the journalese of the LOB-FLOB and Brown-Frown pairs of corpora. At the same time, they found a decline in their use in the academic genre, especially in the American variety of English. At a later stage in this subchapter I will also discuss the diachronic evolution of the continuous VPCs with *up* in the newspaper section of COHA, but for the time being, it is useful to take a look at the synchronic distribution of these constructions in the subgenres of the newspaper section of COCA. As Figure 61 shows, the continuous VPCs with *up* are used most frequently in the financial articles of newspapers, with a normalized frequency of 1,611.63, followed, in order, by lifestyle articles, sports material, an ambiguous miscellaneous category, then, regional and local newspapers, national and international news, and finally, editorials.

Figure 61. Distribution of the continuous VPCs with *up* in the Newspaper subcategories of COCA

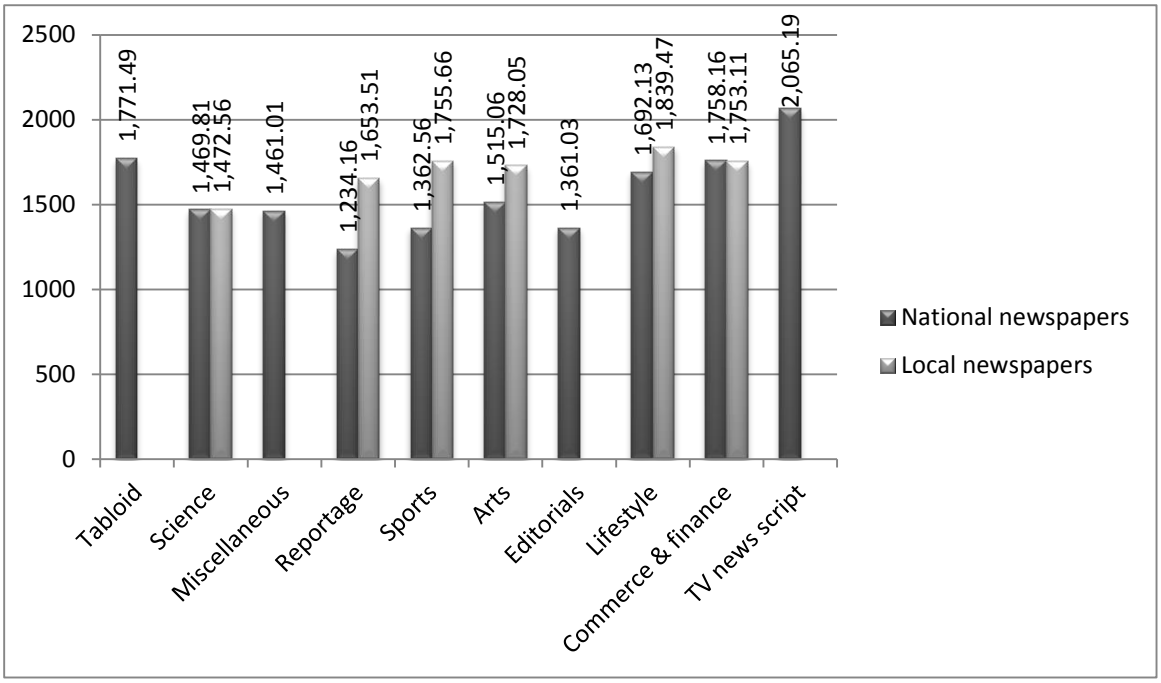


The chart makes sense, as the writing styles employed in lifestyle, sports and financial articles are likely to be less formal, while articles introducing news, and opinion articles written by the senior editorial staff or publisher of a newspaper, are more likely to use a formal style. In fact, the choice of register and style in the press is much more complex than one might think. In a study on the language used in the press, approached from a critical linguistics perspective, Fowler (1991) argues that news is a constructive practice, i.e. the news does not neutrally reflect social reality and empirical facts, but instead, it constructs a social reality by means of various journalistic techniques, with a specific goal. One such technique consists of making the printed medium suggest conversation (cf. Fowler 1991: 61). Fowler explains that, as a person, a reader's communication mode is speech, while the communication mode of a newspaper, as an institution, is print. From a sociocultural point of view, print is associated with formality and authority, while speech suggests informality and solidarity. These basic differences between the two communication modes form a communicative gap between the newspaper and the reader, and the goal of the technique mentioned above is to narrow this gap (cf. Fowler 1991: 59). Fowler points out that the range of linguistic features which contribute to the illusion of conversational style in print media is wide, and encompasses typographic and orthographic features, register, syntax, morphology, and deixis (cf. Fowler 1991: 63). In terms of orthography, newspapers quite often resort to deliberate misspellings and simplifications of spellings in order to suggest orality. In terms of register, the lexis and

vocabulary used in newspapers tend to be deliberately colloquial, in order to mimic the oral mode. Fowler points out that slang, idioms, clichés, proverbs and catch-words are intentionally employed by the popular and tabloid press, while cultivated words are carefully avoided, unless they are scoffed. Similarly, first names, diminutives, and nicknames are used to suggest informality, intimacy, or face-to-face discourse (cf. Fowler 1991: 63). In terms of syntax and morphology, contractions of auxiliaries and negatives (as already discussed in this thesis, in the chapter devoted to contractions), elisions, short and incomplete sentences are all used in newspapers with the same purpose of creating the illusion of oral mode (cf. Fowler 1991: 63). Fowler points out that deictic elements, which are characteristic of speech, such as first and second person personal pronouns, i.e. the ones referring to the speaker and listener, used either separately, or together (*we* with its 'inclusive' meaning, i.e. *you and I*), indicators of time such as *today, now* and *then*, indicators of place, such as *here* and *there*, and the demonstratives *this* and *that*, are used in the newspapers to suggest the oral mode (cf. Fowler 1991: 63).

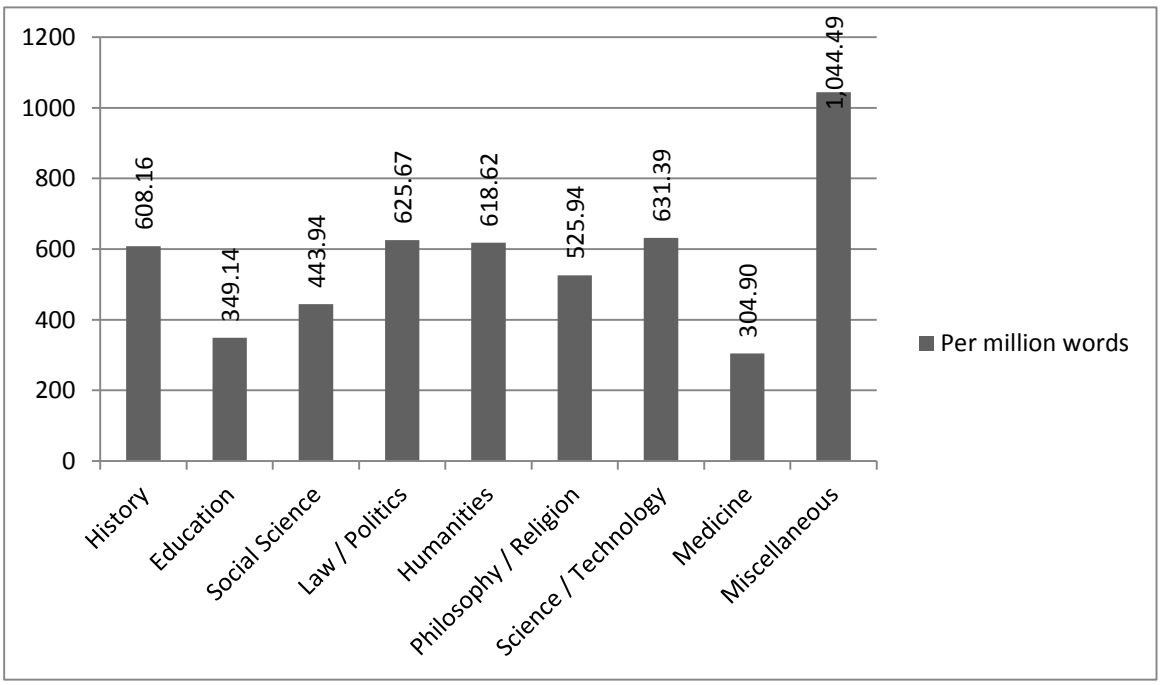
As far as British English is concerned, the more prominent presence of continuous VPCs with *up* in the regional and local newspapers than in the national press is noteworthy (Figure 62). Except for the science and the commerce & finance material, where the normalized frequencies are roughly similar in both national and local press, all other newspaper subgenres exhibit a higher frequency in the regional and local newspapers than in the national broadsheets. The tabloid, miscellaneous, editorial, and television news script categories do not have regional and local counterparts in the BNC. Regarding the nationwide press, the chart is topped by the TV autocue data, with a frequency per million words of 2,065.15, followed, in order, by the tabloid press, with a normalized frequency of 1,771.45, the commerce & finance subcategory, lifestyle, arts, science, miscellaneous, sports, editorials, and finally, the home and foreign news reportage. It is surprising that it is the TV news script category that holds the highest frequency of VPCs, and not the tabloid press, as one might have expected. Surprising also is the higher frequency of continuous VPCs in arts than in the sports and reportage categories. As for the regional and local newspapers, the highest frequency is found in the lifestyle material, closely followed by sports, commerce & finance, arts, reportage, and science.

Figure 62. Distribution of the continuous VPCs with *up* in the Newspaper subcategories of BNC



Within the academic prose category in COCA, the highest normalized frequency is found in the subcategory labeled 'miscellaneous', followed by science and technology, law and politics, humanities, history, philosophy and religion, social science, education, and medicine (Figure 63).

Figure 63. Distribution of the continuous VPCs with *up* in the Academic subcategories of COCA

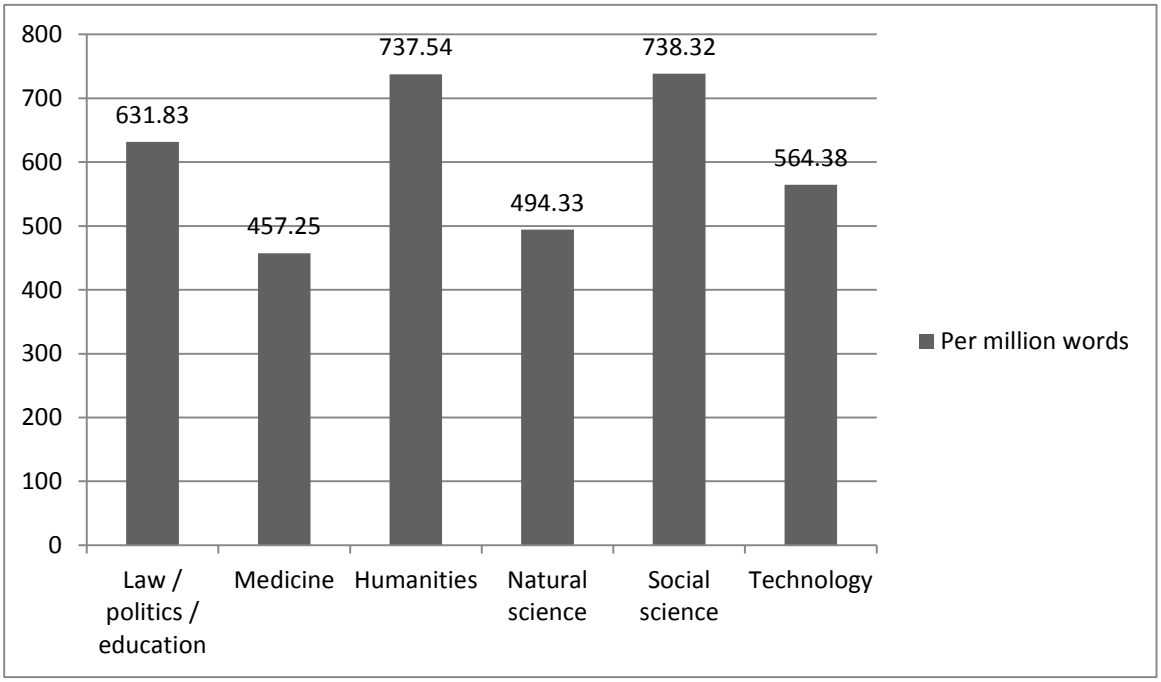


The low frequencies in medicine, education, and social science are as expected, since these fields are inclined towards a more formal style, and so is the high frequency found in science and technology, particularly the latter, as this domain frequently makes use of its own jargon, quite often informal, scattered with idiomatic expressions, e.g. *boot up* ('to start up the operating system of a computer'). Conversely, the relatively high frequency in philosophy and religion is surprising, as these are expected to embrace a highly formal style. In (105) below, extracted from the science & technology subcategory of the academic section of COCA, the author preferred the idiomatic VPC *dig up* instead of the more formal choices *exhume* or *unearth*:

(105) To vampire killers who are **digging up** a corpse, anything unexpected is taken for evidence of vampirism. (COCA)

In the BNC, the distribution of continuous VPCs with *up* in the academic field is highest in social science, with a normalized frequency of 738.32, very closely followed by the humanities, with 737.54; law, politics & education, with 631.83; technology, natural science and medicine (Figure 64).

Figure 64. Distribution of the continuous VPCs with *up* in the Academic subcategories of BNC



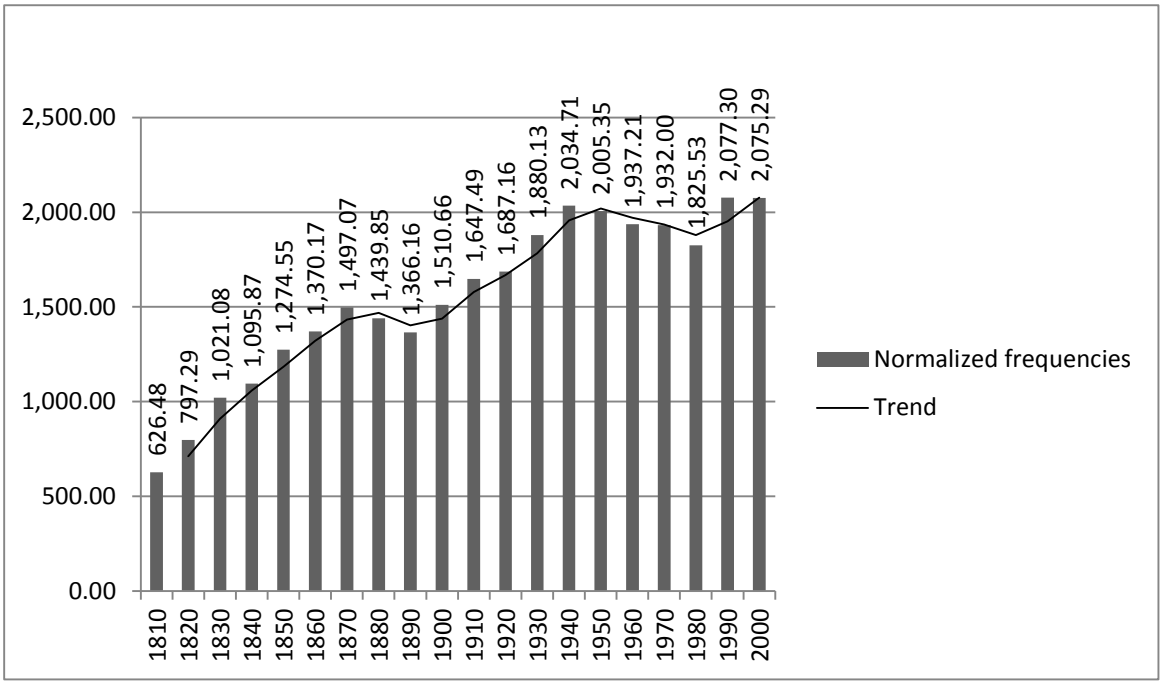
The different sub-categorization of the academic genre in the two corpora makes comparison difficult but not impossible. For instance, philosophy and religion, which are humanistic

academic disciplines, are treated separately by COCA. Similarly, history, which can be regarded as belonging to the humanities, but at the same time to the social sciences, too, is also treated separately by COCA. On the other hand, in the BNC, education is embedded in the law & politics subcategory, while COCA has two separate subcategories, one for law & politics, and another for education. Moreover, COCA merges science and technology in a single subcategory, while the BNC has separate categories for each of them, labeled 'natural science' and 'technology', respectively. Lastly, there is the ambiguous 'miscellaneous' category in COCA. However, a number of remarks can be made despite the different structures of the two corpora. Firstly, medicine, which is independently represented in both corpora, has the lowest frequency of continuous VPCs with *up* in both British and American English. The usage of this type of combinations is roughly one third higher in the former variety. Secondly, if we merge the natural science and technology subgroups in BNC into a single category by adding up the two values and then dividing the total thus obtained by two, we get an average normalized frequency of 529.35, and we can now compare the newly obtained category to the 'science and technology' subgroup in COCA. Thus, we notice a more widespread use of continuous VPCs with *up* in American English than in British English in the academic disciplines of science and technology. Similarly, if we merge education with law & politics in COCA, by the same simple calculation, we get a comparable average normalized frequency of 487.40, which is lower than the frequency of 631.83 found in the BNC.

Except for two short descending curves, the first between the decade of the 1870s and 1890s, and the second between the 1940s and 1980s, the diachronic evolution of the continuous VPCs with *up* in American English has been markedly increasing since 1810 up to the present, from a normalized frequency of as low as 626.48 in the 1810s to a stunning figure of 2,075.29 in the 2000s, i.e. the average frequency has more than tripled (see Figure 65). Below is an excerpt from Henry James's 1875 novel *Roderick Hudson*, retrieved from the fiction section of COHA, from a decade (the 1870s) when the first peak of development of continuous VPCs with *up* seems to have been reached:

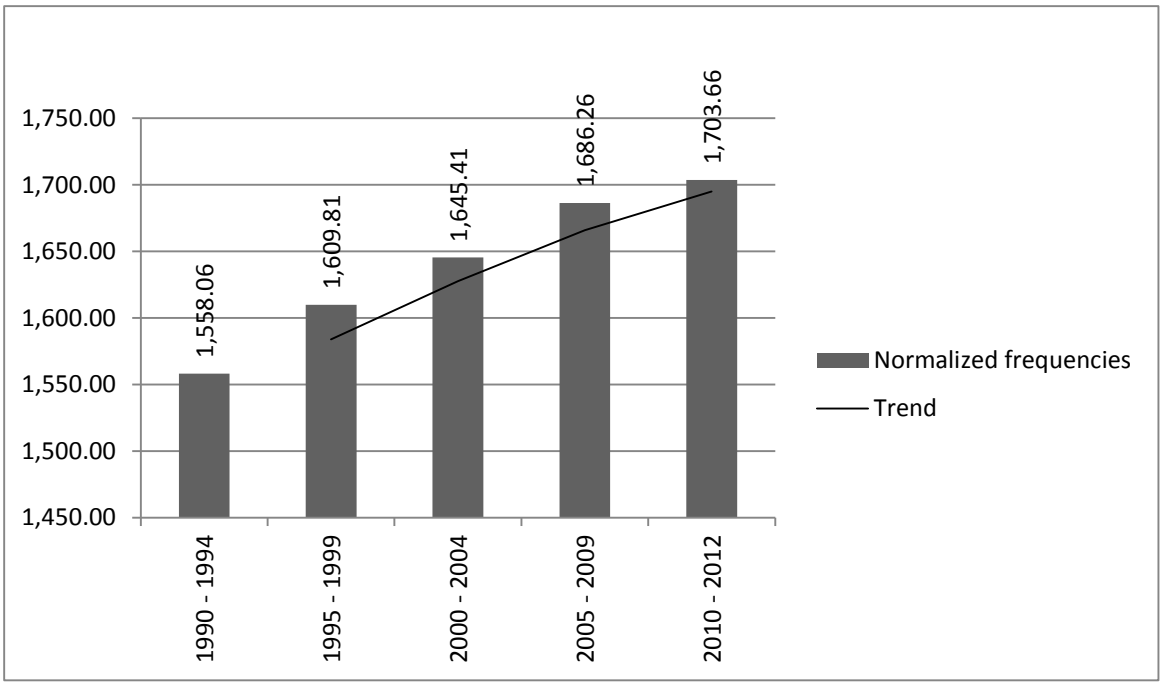
(106) He **sprang up** and stood looking after her until she rounded a turn in the avenue. (COCA)

Figure 65. Diachronic view of the continuous VPCs with *up* in COHA



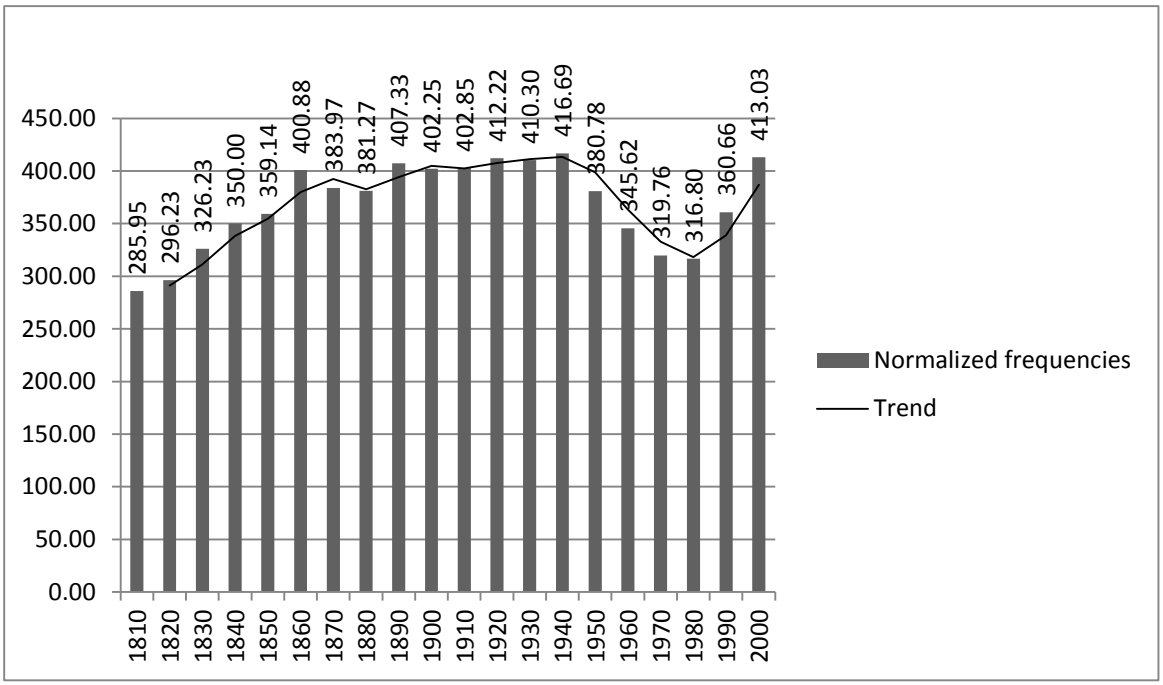
The increase in frequency in American English is evident over even a much shorter and more recent period of time, namely that between 1990 and 2012, as a COCA query reveals (Figure 66).

Figure 66. Diachronic view of the continuous VPCs with *up* in COCA (1990 – 2012)



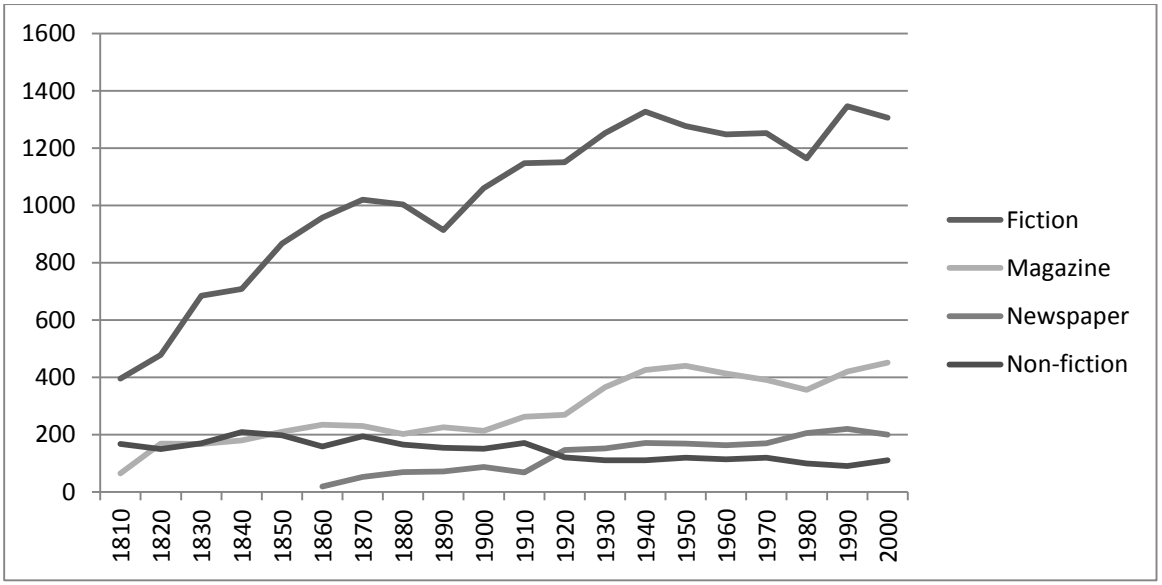
These data are also corroborated by a query in Google Books American English corpus, where, despite a seeming stagnation between the 1890s and the 1940s, and quite a marked decline since the 1940s until the 1980s, the general trend is ascending, from a frequency per million words of 285.95 in 1810, to 413.03 in the 2000s, as Figure 67 shows.

Figure 67. Diachronic view of the continuous VPCs with *up* in Google Books American English



In terms of genre, fiction has seen the most marked growth in COHA, followed by the magazine category, and by newspapers (Figure 68).

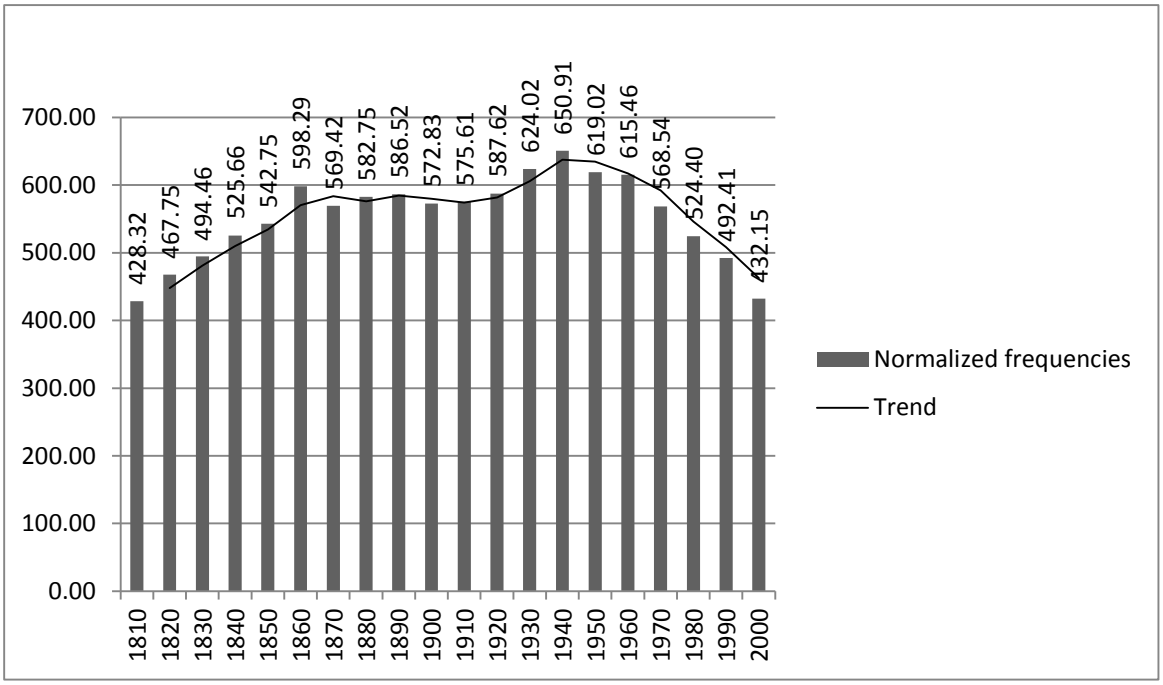
Figure 68. Evolution of the continuous VPCs with *up* in COHA by genre



Note that the earliest available data for the newspaper subgenre in COHA date from the 1860s, which, of course, does not mean that verb-particle constructions had not been used in newspapers before that decade. Contrary to all the other genres, non-fiction has seen a general slightly descending trend, with an apparently modest revival since the 1990s.

Interestingly, the results retrieved from Google Books British English corpus suggest a decline in the use of continuous VPCs with *up* in British English since the 1940s up to the present, after an increase between the 1810s and the 1940s (Figure 69). However, if we compare the frequency of the 2000s, of 432.15, to the one of the 1810s, of 428.32, we can still notice a very slight growth.

Figure 69. Diachronic view of the continuous VPCs with *up* in Google Books British English



5.4.2.2. Discontinuous VPCs with *up*

The search for discontinuous verb-particle constructions in the corpora is quite complex, due to the various items which can intervene between the verb and the particle. One of the commonest possibilities is a noun, usually preceded by the definite article *the*, as in the example below, extracted from the newspaper section of COCA:

(107) **Turn the heat up** and add 1 cup of red wine, and cook until the pan is nearly dry, stirring constantly. (COCA)

There are however instances where the noun is preceded by the indefinite article *a* or *an*:

(108) Because not everybody embraces the beauty of Jason Statham **beating a guy up** while strapped to a chair and then jumping out a three-story window, Ruthe Stein has also picked several promising independent films as well. (COCA)

In some cases the noun is preceded by a determiner, as in (109) below:

(109) "I think the government has to take a large share of the blame for **stirring these divisions up**, but the concern is once they **stir these things up**, it's not easy to control them, it's not easy to put them back in the box," says Ms. Kinninmont. (COCA)

In some other cases the determiner alone is inserted between the verb and the particle:

(110) **Serve this up** with a teaspoon, or make your own dressing by drizzling on a little olive oil and as much red wine vinegar as you'd like. (COCA)

Last but not least, in some cases, the item which intervenes between the verb and the particle is a pronoun, as in the example below extracted from the magazine section of COCA:

(111) At 11 p.m., a low rumbling noise **woke me up** from a restless sleep. (COCA)

So many different possible combinations require an equal number of different search strings, which for convenience are listed in Table 13. Of course, many other variants are possible, as for instance, in (112) below, where along with the noun and the definite article a modifier has also been included between the verb and the particle. Nevertheless, I decided to limit myself to the four common combinations listed in Table 13.

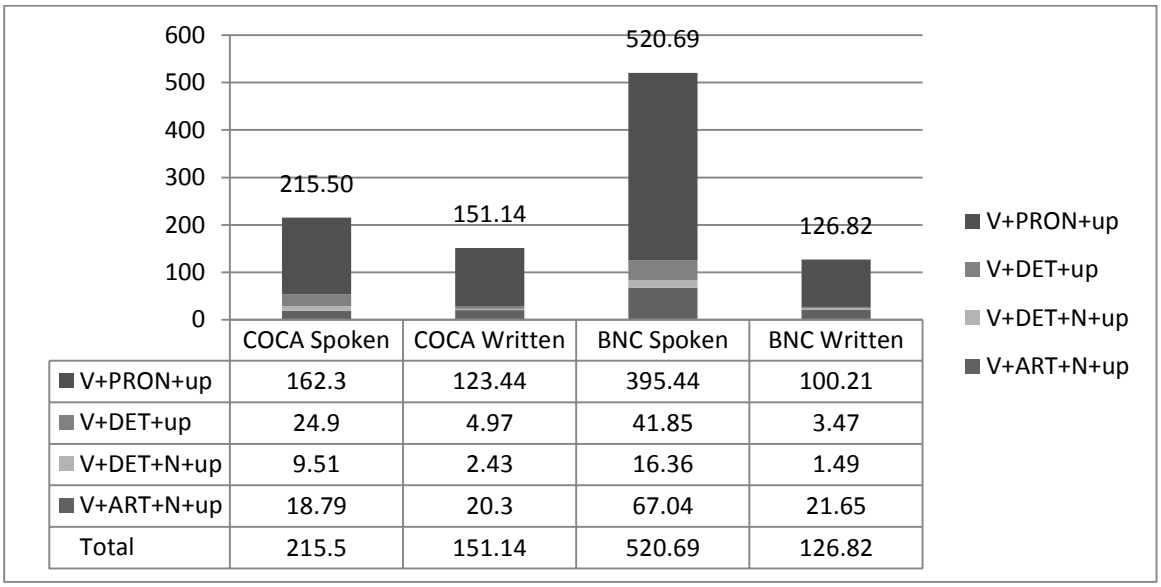
(112) Then we began working with a four-track machine. Wow. We could then do double vocals, and it **thickened the sound right up**. (COCA)

A first noteworthy observation which can be made with respect to the use of discontinuous VPCs with *up* is that the constructions with a pronominal direct object (DO) are by far the most frequent in both British and American English, in both speech and writing (Figure 70). As the corpus evidence shows, the second commonest type of combinations is that with a

nominal DO, usually preceded by the definite article. The constructions in which the item placed between the verb and the particle is either a noun preceded by a determiner or simply a determiner are not so common in either variety of English, as the data in the table of Figure 70 show. The second striking observation is the significant difference between the total frequency of discontinuous VPCs with *up* in the spoken section of the BNC and the total frequency of the combinations in the corresponding section of COCA, with a much higher frequency in the former. This difference is consistent with the corpus findings for the continuous VPCs with *up*, where the frequency in the BNC is also higher than that in COCA (compare Figure 56 to Figure 70). However, while in the case of the continuous VPCs the frequency in BNC is roughly 11% higher than in COCA, in the case of the discontinuous combinations, the frequency in British English is more than double the frequency in American English. Such a discrepancy can be explained either by a more pronounced preference among British English speakers for verb-particle constructions, or by the different structures of the two corpora.

The third striking aspect relates to the substantially higher frequencies of continuous VPCs in both corpora, in both speech and writing, than the frequencies of discontinuous combinations. If the frequency of continuous VPCs with *up* in the BNC spoken material is almost four times higher than the frequency of discontinuous constructions in the corresponding section of the same corpus, in the case of the written material, the frequency of continuous VPCs is almost 12 times higher.

Figure 70. Frequencies per million words of discontinuous VPCs with *up* in COCA and BNC



	COCA Spoken	COCA Written	BNC Spoken	BNC Written
V+PRON+up	162.3	123.44	395.44	100.21
V+DET+up	24.9	4.97	41.85	3.47
V+DET+N+up	9.51	2.43	16.36	1.49
V+ART+N+up	18.79	20.3	67.04	21.65
Total	215.5	151.14	520.69	126.82

In terms of genre, fiction has the highest frequency per million words in COCA, with a total of 352.16 discontinuous VPCs with *up*, followed by magazine, with a normalized frequency of 130.27, newspapers, with 95.29, and the academic category, with 26.85 (Table 5). The distribution in the BNC is similar to that in COCA, with the highest total normalized frequency in fiction, followed by magazines, newspapers and academic prose. These distributions of discontinuous VPCs in the two corpora are consistent with the distribution of continuous VPCs presented in Figure 57. The only exception is the slightly higher frequency of continuous VPCs in the newspaper category of the BNC than that found in the magazine genre of the same corpus. However, all the frequencies of the continuous VPCs with *up* in the genres of both corpora are considerably higher than the frequencies of the discontinuous combinations found in the two corpora.

In terms of DO type, the discontinuous constructions with a pronoun between the verb and the particle have the highest frequencies in all four genres of both corpora. They are followed, with considerably lower frequencies, by the combinations with a nominal DO preceded by an article, the constructions with a determiner between the verb and the particle, and finally, the construction with a nominal DO preceded by a determiner.

Table 5. Normalized frequencies of discontinuous VPCs with *up* in the written material of COCA and BNC, divided by genre

	Fiction		Magazine		Newspaper		Academic	
	COCA	**BNC**	**COCA**	**BNC**	**COCA**	**BNC**	**COCA**	**BNC**
V+Art+N+*up*	44.57	43.81	18.88	25.75	13.25	21.12	4.50	6.72
V+Det+N+*up*	4.24	3.27	2.53	1.93	2.29	1.05	0.66	0.52
V+Det+*up*	11.42	7.61	3.91	3.99	3.78	2.39	0.77	1.96
V+PRON+*up*	291.93	301.33	104.95	90.88	75.97	77.87	20.92	18.07
TOTAL	**352.16**	**356.02**	**130.27**	**122.55**	**95.29**	**102.43**	**26.85**	**27.27**

Below are a few examples extracted from the magazine section of COCA, one for each type of the mentioned discontinuous combinations.

(113) Instead of **blowing the place up**, Jennifer Bright transformed it into a home that blows folks away. (COCA; magazine; V + ART + N + *up*)

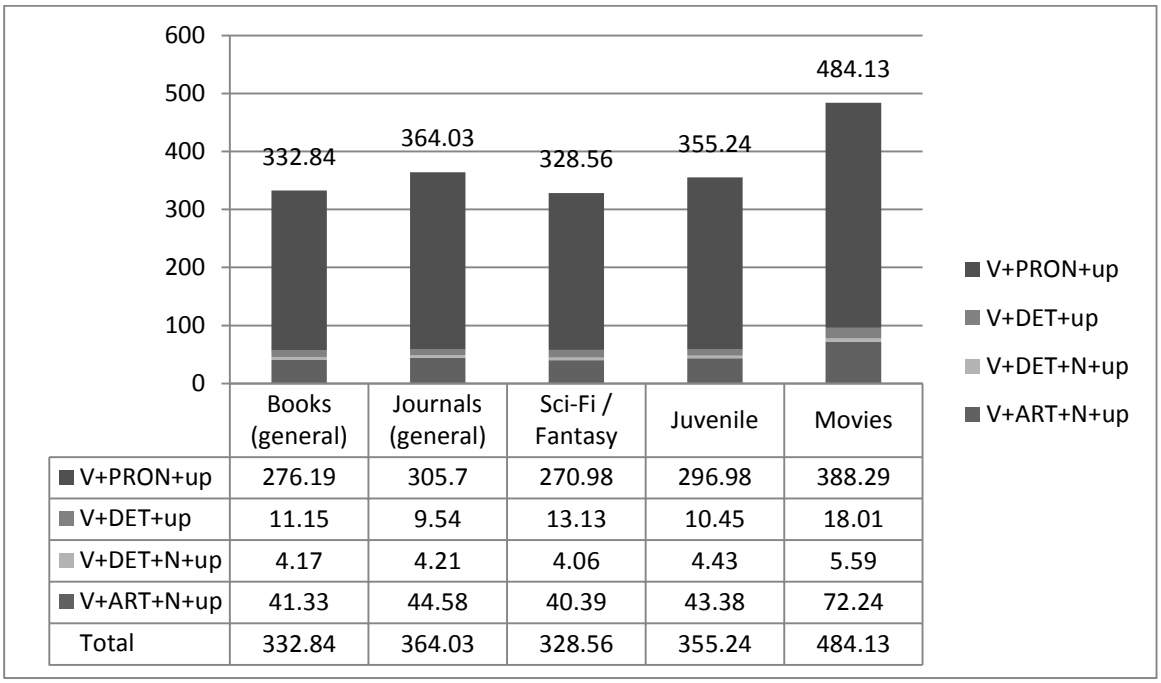
(114) Israel may not **give this land up**, now or ever. (COCA; magazine; V + DET + N + *up*)

(115) Agudelo played a confident and menacing 90 minutes, then **followed that up** with an assist on New York's only goal in his second professional start. (COCA; magazine; V + DET + *up*)

(116) The billions of dollars for incarceration would be better spent on sending young people to good schools and colleges, rather than **locking them up** at early ages. (COCA; magazine; V + PRON + *up*)

Except for fiction, where the total normalized frequency is slightly higher in the BNC than in COCA, in all other three written genres, the frequencies are higher in American English than in the British variety. A more detailed look into fiction reveals the distribution of the discontinuous VPCs with *up* in the subcategories of this genre. Thus, movies have the highest total normalized frequency, namely 484.13 (Figure 71). They are followed by journals targeted at a general readership, with a total normalized frequency of 364.03; youth fiction, with 355.24; books for the general public, with 332.84; and science fiction and fantasy literature with a frequency per million words of 328.56. The commonest item which intervenes between the verb and the particle is a pronoun, as the high frequencies of this type of combination in all subgenres of fiction show in Figure 71. The second commonest combination type in terms of DO is that with a nominal DO preceded by an article, which is usually, but not necessarily always, the definite article. The discontinuous combinations in which what intervenes between the verb and the particle is either a determiner alone or a noun preceded by a determiner have significantly lower frequencies in all fiction subgenres. The distribution of the discontinuous VPCs in the subcategories of fiction is roughly similar to that of the continuous constructions in fiction, with the only difference that the frequency in the juvenile literature is slightly higher than that in journals in the case of continuous combinations.

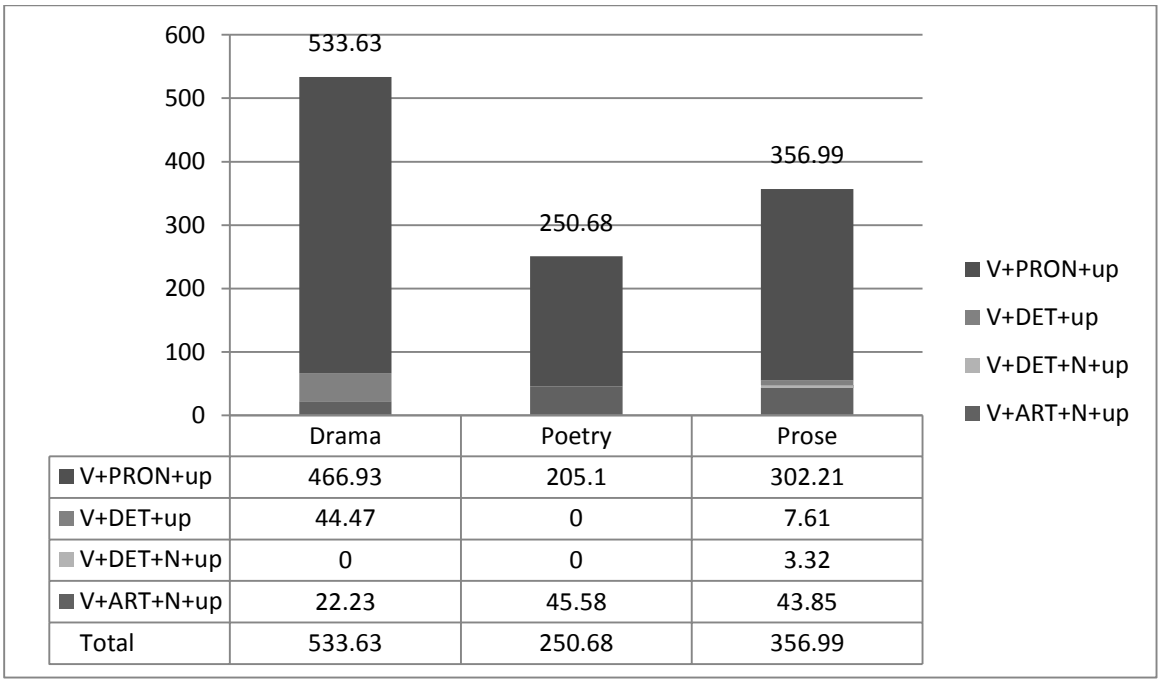
Figure 71. Normalized frequencies of the discontinuous VPCs with *up* in the Fiction subgenres of COCA



	Books (general)	Journals (general)	Sci-Fi / Fantasy	Juvenile	Movies
V+PRON+up	276.19	305.7	270.98	296.98	388.29
V+DET+up	11.15	9.54	13.13	10.45	18.01
V+DET+N+up	4.17	4.21	4.06	4.43	5.59
V+ART+N+up	41.33	44.58	40.39	43.38	72.24
Total	332.84	364.03	328.56	355.24	484.13

Of the three fiction subgenres of the BNC, drama has the highest total normalized frequency of discontinuous VPCs with *up*, followed by prose, and then poetry, a hierarchy which reflects the identical distribution of the continuous combinations, as may be noticed from a comparison of Figure 72 to Figure 59.

Figure 72. Normalized frequencies of the discontinuous VPCs with *up* in the Fiction subgenres of BNC



	Drama	Poetry	Prose
V+PRON+up	466.93	205.1	302.21
V+DET+up	44.47	0	7.61
V+DET+N+up	0	0	3.32
V+ART+N+up	22.23	45.58	43.85
Total	533.63	250.68	356.99

However, the frequencies of the continuous VPCs are, as has been noticed before, substantially higher than those of discontinuous combinations. While the constructions in which the item that intervenes between the verb and the particle is a pronoun have, as usual, the highest frequencies in all three fiction subgenres of the BNC, no matches of combinations with a determiner and with a nominal DO preceded by determiner were found in poetry. The same is true for the combinations with determiner plus noun in drama. Contrary to the general tendency in the fiction subcategories of COCA and in the prose subgenre of the BNC, where the frequencies of combinations with a nominal DO are higher than those of VPCs with only a determiner, the situation in the fiction subcategory of drama in the BNC is reversed, that is, the frequency of the constructions with a determiner alone between the verb and the particle is double the frequency of discontinuous VPCs with a nominal DO. Another interesting fact is the higher frequency per million words of discontinuous combinations with a nominal DO in poetry than in drama and prose. Below is an example of such a discontinuous VPC with *up*, extracted from the poetry subgenre of the fiction category of the BNC:

(117) The sonar arm # can **pick the shoal up** as an entity, then between the boats # the net falls open like a mouth and at its lips # the bright pelagic fish acquiesce. (BNC: FBG)

However the number of tokens in poetry is only 10, compared to 686 in prose.

A look at the distribution of discontinuous VPCs with *up* in the magazine subgenres of COCA shows that the highest total frequency per million words is found in women's and men's magazines, followed by children magazines, entertainment, sports, African-American, health, news, science and technology magazines, financial, society and arts, and finally, religious magazines (Table 6). As usual, the frequencies of the continuous constructions are significantly higher than those of the discontinuous combinations. Both types of constructions have the highest frequencies in women's and men's magazines, and they are the least common in religious magazines. Otherwise, the distributions of the two types of constructions in the magazine subgenres of COCA are quite different, as a comparison between Table 6 and Figure 60 shows. The bulk of the discontinuous VPCs with *up* in the magazine subgenres of COCA is of the V + PRON + *up* type, and the second commonest item which intervenes between the verb and the particle is a noun phrase with an article. As usual so far, the least frequent combinations are those with a determiner between the verb and *up*, and with a noun plus determiner, respectively.

Table 6. Normalized frequencies of the discontinuous VPCs with *up* in the Magazine subgenres of COCA

	News	Finance	Sci./Tech.	Soc./Arts	Religion	Sports	Entertain.	Health	African	Children	Women/Men
V + ART + N + *up*	12.20	15.03	18.91	13.99	6.79	37.70	22.60	22.24	11.56	27.53	19.07
V + DET + N + *up*	1.72	3.42	2.22	1.18	0.47	4.16	4.42	3.01	3.58	1.84	2.67
V + DET + *up*	3.55	4.57	2.29	2.48	3.51	5.27	5.65	3.38	3.58	1.22	7.49
V + PRON + *up*	82.14	67.72	68.14	58.82	62.75	138.98	161.91	103.89	116.98	177.41	210.20
TOTAL	**99.61**	**90.74**	**91.56**	**76.47**	**73.52**	**186.11**	**194.58**	**132.52**	**135.7**	**208**	**239.43**

The example below, where a discontinuous VPC with a pronominal DO is employed in a children's magazine, is extracted from COCA:

(118) The insect falls into the water, and the fish **gobbles it up**. (COCA)

The verb *gobble* was preferred to the more common *swallow*, while the inference brought about by the addition of the particle *up* is that the action of gobbling is completed.

As already mentioned, the BNC does not provide a classification of its magazine genre.

The most frequent discontinuous VPCs with *up* are found in the sports articles of newspapers in American English, as Table 7 shows. The second highest total normalized frequency is found in lifestyle articles, followed in descending order by articles dealing with regional and local news, the miscellaneous category, editorials, national, and lastly, international news reports. Like in all other genres and subgenres so far, the discontinuous combinations are considerably less frequent than the continuous VPCs in the subcategories of newspapers, and the distributions of the two combination types are quite different, as may be noticed from a comparison of Table 7 to Figure 61. As usual, the commonest constructions are those with a pronominal DO, followed by those with a noun accompanied by an article. Contrary to the generally higher frequencies of the combinations with a determiner as DO than the frequencies of the constructions with a determiner and a noun, the frequencies of the discontinuous VPCs with a noun preceded by a determiner are higher than those of the

constructions with a determiner alone in American English, in national and international news reports, and in financial articles.

Table 7. Normalized frequencies of the discontinuous VPCs with *up* in the Newspaper subgenres of COCA

	Misc.	Intl. news	Nat. news	Local news	Money	Life	Sports	Editorial
V + ART + N + *up*	12.15	7.26	10.33	14.26	12.68	14.80	19.25	11.31
V + DET + N + *up*	2.31	1.87	2.43	2.16	2.92	1.76	2.99	2.04
V + DET + *up*	3.53	1.41	3.34	3.81	2.28	4.22	6.20	3.17
V + PRON + *up*	75.04	48.49	57.57	76.62	46.28	90.99	104.6	76.26
TOTAL	**93.03**	**59.03**	**73.67**	**96.85**	**64.16**	**111.77**	**133.04**	**92.78**

In British English, the highest total normalized frequency is found in the scientific articles of regional and local newspapers, followed in order by the tabloid press, TV autocue data, lifestyle articles published in the broadsheet national newspapers, sports material in regional and local newspapers, the miscellaneous category, regional and local reportage, regional and local lifestyle articles, sports material in the nationwide press, articles related to arts published in broadsheet national newspapers, editorials, financial articles in national press, material related to science published in national newspapers, articles about arts in regional and local newspapers, finance related material in the regional and local press, and lastly, nationwide reportage (see Table 8). As already mentioned, there are no regional and local counterparts for the tabloid, miscellaneous, editorial, and television news script categories in the BNC. There is a discrepancy between the frequency of the scientific material in the regional and local newspapers, which is the highest in the newspaper subgenres of the BNC, and the frequency of the scientific material in the nationwide press, which holds only the thirteenth position in the frequency chart. Except for the tabloid, local reportage, local sports and national arts categories, no matching data were found for the V + DET + N + *up* construction in the other newspaper subgenres of the BNC. As usual, the most frequent discontinuous combination in all subgenres of the newspaper category of the BNC is V + PRON + *up*, followed by the VPCs with a nominal DO preceded by an article, the constructions with a determiner

preceding *up*, and those with a determiner and a noun. Contrary to this general tendency, in the case of the science material published in the broadsheet national newspapers, the normalized frequency of the combinations with a determiner as direct object is 15.47, while no matching data were found for the V + DET + N + *up* combination.

Table 8. Normalized frequencies of the discontinuous VPCs with *up* in the Newspaper subgenres of BNC

	Tabloid	Science		Misc.	Reportage		Sports		Arts		Editorial	Lifestyle		Finance		Script
		Nat.	**Loc.**		**Nat.**	**Loc.**	**Nat.**	**Loc.**	**Nat.**	**Loc.**		**Nat.**	**Loc.**	**Nat.**	**Loc.**	
V + ART + N + *up*	19.62	0	55.22	23.53	12.20	23.12	27.32	15.84	8.67	12.74	9.93	49.41	16.88	40.83	24.55	22.97
V + DET + N + *up*	2.80	0	0	0	0	2.61	0	0.99	2.89	0	0	0	0	0	0	0
V + DET + *up*	5.61	15.47	0	0.98	1.53	2.61	3.41	3.96	0	0	0	0	0	2.40	2.46	3.17
V + PRON + *up*	98.10	61.89	73.63	84.33	36.61	77.94	58.05	99.02	72.28	50.95	69.54	74.11	88.86	36.03	31.92	97.44
TOTAL	**126.13**	**77.36**	**128.85**	**108.84**	**50.34**	**106.28**	**88.78**	**119.81**	**83.84**	**63.69**	**79.47**	**123.52**	**105.74**	**79.26**	**58.93**	**123.58**

That academic prose is the most reluctant genre to the informal style is confirmed once again by the low frequencies of discontinuous VPCs with *up* in the academic subcategories of COCA. The highest total normalized frequency is in miscellaneous with 77.05, followed by the humanities, with a total frequency per million words of 37.57; philosophy and religion, with 29.23; law and politics, with 26.63; science and technology, with 23.23; social science, with a normalized frequency of 22.25; history, with 20.41; education, with 15.25; and finally, medicine, with a frequency of as low as 12.4 (Table 9). In terms of DO type, pronouns are, again, the most frequent, followed by nouns preceded by an article, then a determiner alone, and the least common, determiner plus noun. The only two subcategories where the V + DET + N + *up* construction is more frequent than the VPCs with only a determiner as DO are social sciences, and philosophy and religion, while in education and humanities, the two combinations have very similar normalized frequencies of 0.53 and 0.59, respectively.

Table 9. Normalized frequencies of the discontinuous VPCs with *up* in the Academic subgenres of COCA

	History	Education	Soc. Sci.	Law/Politics	Humanities	Phil./Religion	Sci./Tech.	Medicine	Misc.
V + ART + N + *up*	3.02	2.97	3.65	3.95	6.79	4.01	5.47	2.84	8.46
V + DET + N + *up*	0.65	0.53	0.56	0.58	0.59	1.48	0.64	0.30	1.17
V + DET + *up*	0.73	0.53	0.49	1.05	0.59	0.74	0.85	0.60	1.64
V + PRON + *up*	16.01	11.22	17.55	21.05	29.60	23.00	16.27	8.66	65.78
TOTAL	**20.41**	**15.25**	**22.25**	**26.63**	**37.57**	**29.23**	**23.23**	**12.4**	**77.05**

In (119) below, extracted from a theological study found in COCA, the use of the phrasal verb *hold up* was preferred to the more formal alternative *raise*.

(119) The curate attending her **held a crucifix up** before her eyes, urging her to gaze upon it. (COCA)

In the BNC, the highest total normalized frequency of discontinuous VPCs with *up* is found in social science, followed by the humanities, the subgenre of law, politics and education, natural sciences, medicine, and finally, technology (see Table 10). If, for comparative purposes, we merge the natural science and technology subgenres into a single group, we get an almost identical frequency of 23.33 to that of the science and technology subgenre in the academic category of COCA. On the other hand, if, for the same comparative purpose, we combine the law and politics subgroup in COCA with education, we obtain a normalized frequency of 41.88, which is almost double the frequency in the corresponding law, politics and education subgroup of the BNC. As usual, the frequencies in the academic subgenres of both COCA and BNC are much higher for the continuous than for the discontinuous constructions. Typically, the pronoun is the item which intervenes the most frequently between the verb and *up*. The second most common discontinuous construction in the academic subgenres of the BNC is V + ART + N + *up*, followed by the combinations with a determiner as DO, and those with a determiner and a noun. In law, politics and education,

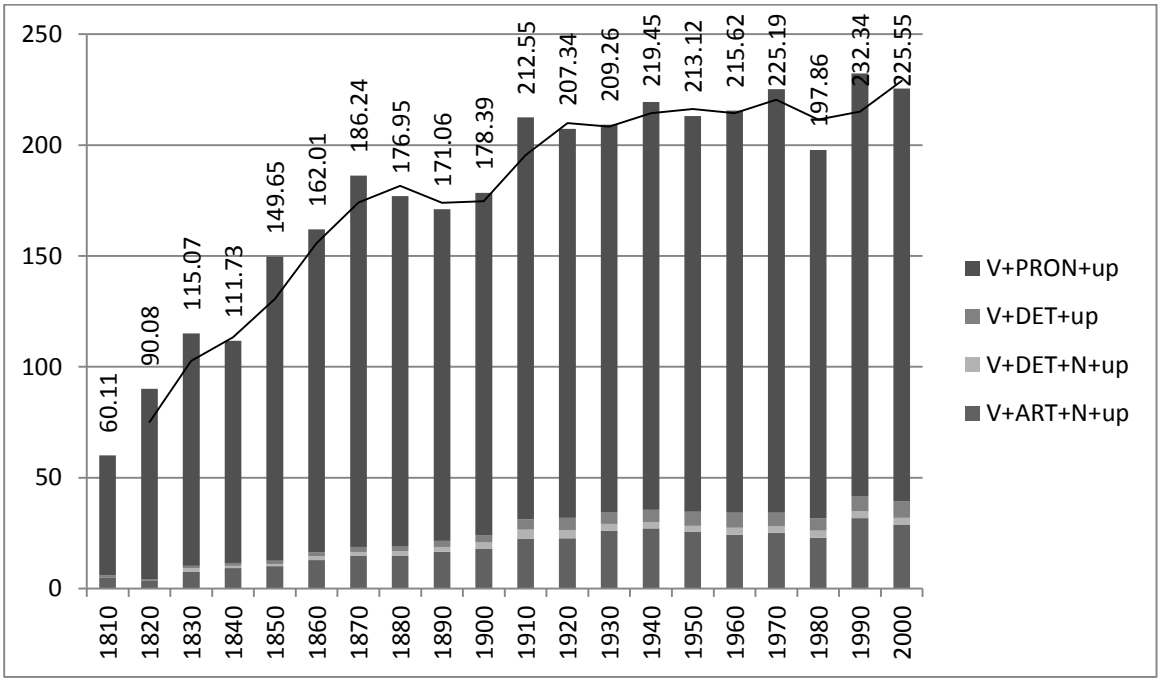
medicine, and technology, no matching data were found for the V + DET + N + *up* construction.

Table 10. Normalized frequencies of the discontinuous VPCs with *up* in the Academic subgenres of BNC

	Law/Pol./Edu.	Medicine	Humanities	Natural science	Social science	Technology
V + ART + N + *up*	5.85	7.08	5.16	6.34	9.23	4.42
V + DET + N + *up*	0	0	0.61	0.91	1.18	0
V + DET + *up*	2.17	0.71	2.12	1.81	2.13	1.47
V + PRON + *up*	14.95	5.66	24.57	5.43	26.28	2.95
TOTAL	**22.97**	**13.45**	**32.46**	**14.49**	**38.82**	**8.84**

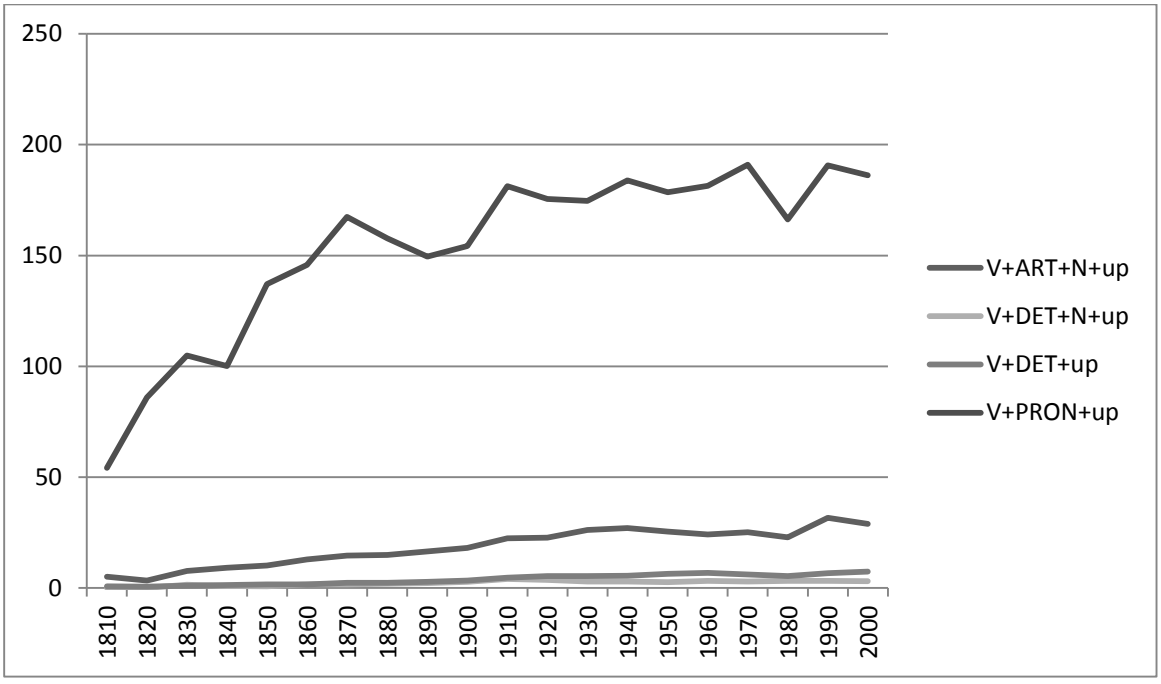
Like the continuous combinations, the discontinuous VPCs with *up* have seen a marked increasing diachronic evolution in American English, from a total normalized frequency as low as 60.11 in the 1810s, to 225.55 in the 2000s, as Figure 73 shows.

Figure 73. Diachronic view of the discontinuous VPCs with *up* in COHA



In fact, the trend line of the discontinuous VPCs is almost identical to that of the continuous combinations, with two short observable periods of decline, between the 1870s and the 1890s, and since the 1970s to the 1980s, respectively. However, the frequencies of the continuous constructions have been, at all times, considerably higher than those of the discontinuous combinations. As in the case of the continuous combinations, the latest tendency of the discontinuous VPCs, too, starting since the 1990s up to the present seems to have been slightly decreased, but this tendency is hardly noticeable, and therefore insignificant, considering the overall substantial growth. In terms of DO type, the most frequent discontinuous combination in American English has been V + PRON + *up*, with considerably higher growth rates at all times, followed by V + ART + N + *up*, V + DET + *up*, and V + DET + N + *up*, with much more constant, but at the same time, less pronounced development rates (see Figure 74).

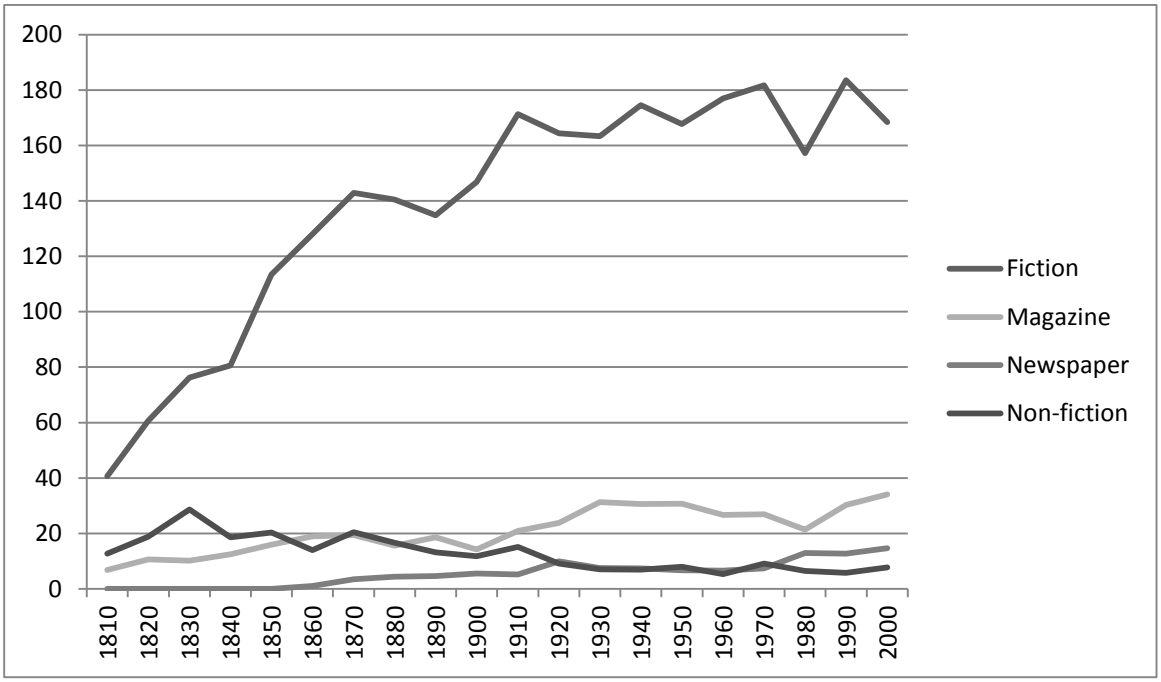
Figure 74. Evolution of the discontinuous VPCs with *up* in COHA by DO type



In order to obtain a comparable frequency chart of the discontinuous VPCs with *up*, I ran all the four searches corresponding to the four discontinuous combination types analyzed in this chapter, for each of the four genres included in COHA, and I used only the totals in the chart. Thus, both the continuous and the discontinuous constructions seem to have had similar developments in American English, as a comparison of Figure 75 to Figure 68 reveals, the only difference between the two being the substantially higher frequencies in the case of the continuous combinations. Fiction has had the highest growth rates at all times, followed by

magazines and newspapers with much milder growing trend lines. The only genre which seems to have seen a decline in the use of the discontinuous VPCs with *up* in American English is non-fiction, which has started from a higher usage rate than both the magazine and newspaper categories back in the 1810s, but has decreased to a lower frequency than the two in the 2000s. Since the term 'non-fiction' used by the creators of COHA may refer to many genres, among which academic texts also, I may consider this decreasing trend a corroboration of Hundt and Mair's (1999) finding of a decreasing usage of phrasal verbs in academic prose in the Brown and Frown corpora, that is, in American English. However, as with the magazine and newspaper genres, the use of both continuous and discontinuous combinations in non-fiction has also seen a seemingly slight revival since the 1990s, as Figure 75 and Figure 68 show.

Figure 75. Evolution of the discontinuous VPCs with *up* in COHA by genre

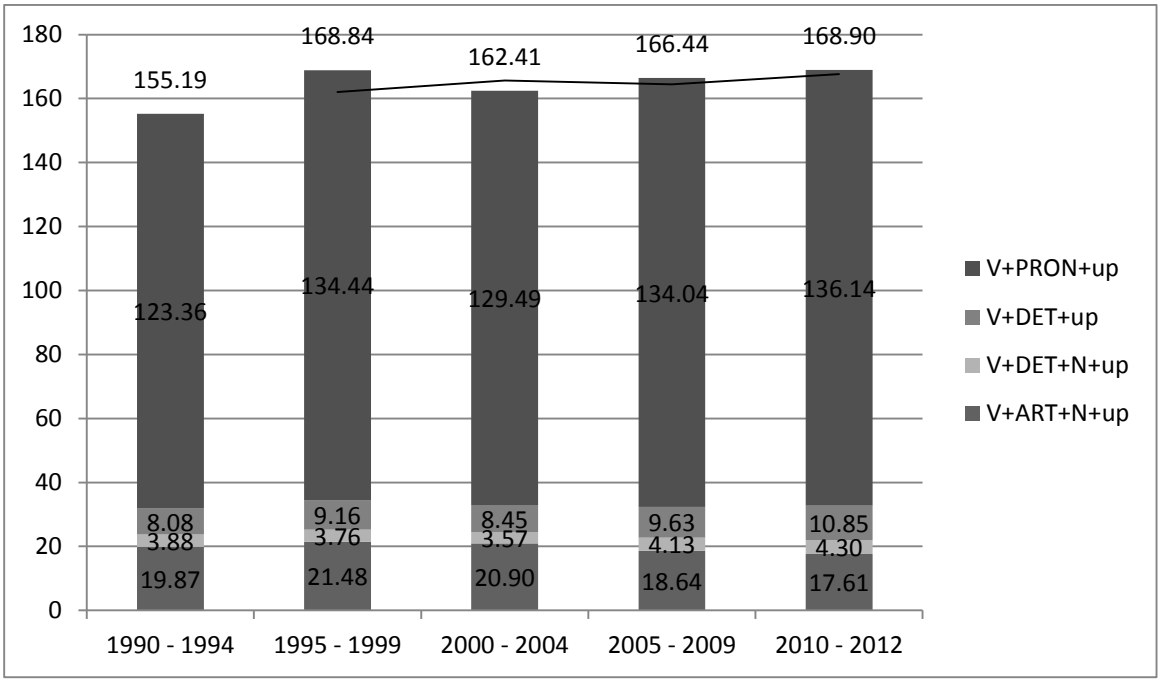


Below is an example of discontinuous VPC with *up* used in a non-fiction book, extracted from the 1970s data of COHA.

(120) Hot stages are also available which are able to **heat the specimen up** to about 800° C; similarly, cold stages are able to cool the specimen to liquid nitrogen temperature. (COHA)

Unlike the continuous combinations, whose growing use in American English is evident even over the last two decades, the development of the discontinuous VPCs with *up* over the same period is not so obvious, as Figure 76 shows. Even if the trend line indicates almost a stagnation of usage, if we compare the frequency per million words of the period between 1990 and 1994 to that of the period between 2010 and 2012, we can still conclude that the tendency is one of growth. The same is valid for the evolution over the last three periods included in COCA, where a slight but continuous growing trend is observable, from a normalized frequency of 162.41 in the period between 2000 and 2004, to 168.90 in the period between 2010 and 2012. As usual, the most frequent item which intervenes between the verb and the particle is a pronoun, followed by a noun preceded by article, a determiner, and a noun preceded by a determiner.

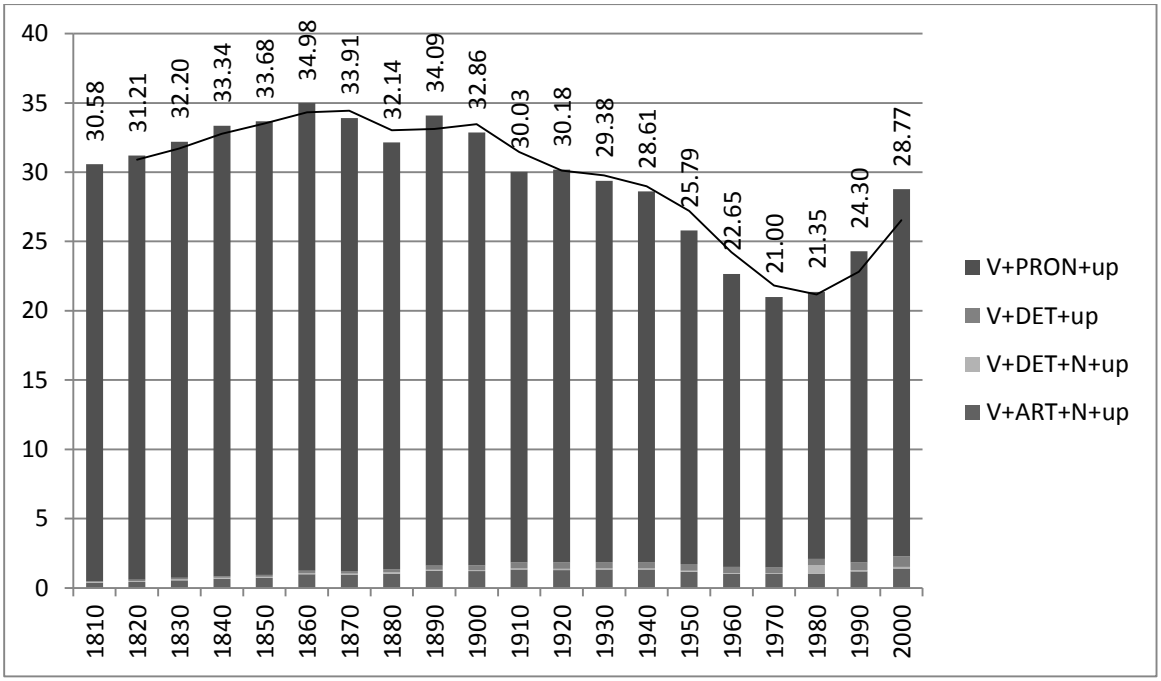
Figure 76. Diachronic view of the discontinuous VPCs with *up* in COCA (1990 – 2012)



The diachronic evolution of the discontinuous VPCs with *up* as suggested by Google Books American English is roughly similar to that of the continuous combinations in the same corpus, with an initial growing trend since the 1810s up to the 1860s, followed by a decline to an all-time low level reached in the 1970s, and followed again, since that point, by a sudden steep revival up to the 2000s (compare Figure 77 to Figure 67). As usual, the overwhelming majority of the discontinuous constructions are those with a pronominal DO, followed by the combinations with a nominal DO preceded by an article. The constructions with a determiner

alone as DO, and those with a determiner and a noun have very low, almost insignificant frequencies per million words.

Figure 77. Diachronic view of the discontinuous VPCs with *up* in Google Books American English

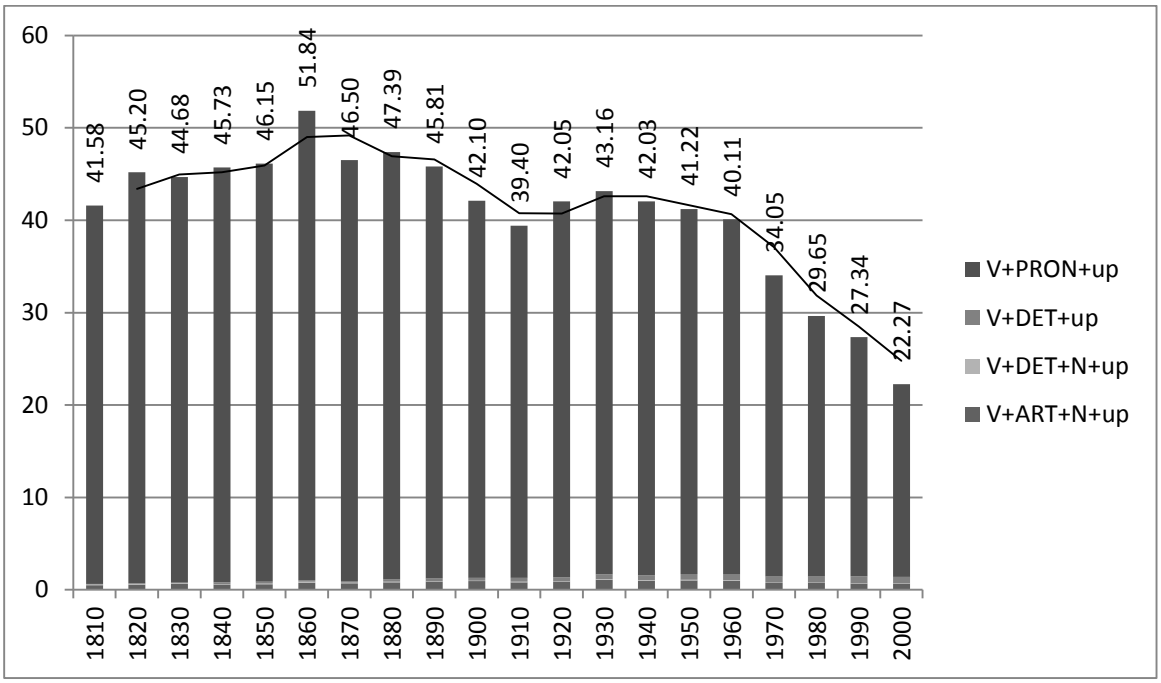


The following example of the use of discontinuous VPC with *up* was extracted from Google Books American English corpus:

(121) An effort was made to **bring the subject up** to the Congress. (Google Books American English Corpus)

The only corpus which indicates a continued decline in the use of the discontinuous VPCs with *up* since the 1930s up to the present is Google Books British English (see Figure 78). This decreasing trend is consistent with the evolution of the continuous constructions in the same corpus (see Figure 69), which might indicate a general decline in the use of phrasal verbs in British English.

Figure 78. Diachronic view of the discontinuous VPCs with *up* in Google Books British English



5.4.3. Conclusions

As the most productive particle in verb-particle constructions, *up* is extensively used to convey a wide range of meanings, from literal to idiomatic and perfective. Of the two syntactic combination types, the continuous one, i.e. the construction in which the particle is adjacent to the verb and the direct object follows the particle, is by far the most commonly used in both British and American English, at the expense of the discontinuous construction, in which the direct object is placed between the verb and the particle. In American English, the growth of phrasal verbs with *up* is obvious, which supports the colloquialization approach in this thesis, and corroborates Hund and Mair's previous corpus-based research on the development of phrasal verbs in English. Unlike American English, the queries performed in Google Books British English corpus indicate a decline in the use of the verb-particle constructions with *up* in British English, since the 1940s up to the present moment. Unlike contractions and the colloquial use of *like*, the difference between the spoken and written language in the use of phrasal verbs with *up* is not so significant, at least in the case of the continuous constructions, which confirms the popularity of phrasal verbs in both speech and writing. In terms of genre, fiction is the most open to the use of phrasal verbs with *up*, in both British and American English, while at the opposite extreme, academic prose has the lowest

frequencies of VPCs with *up* in both varieties of English. Actually, the only genre which in the long term has undergone a decline in the use of phrasal verbs with *up* in American English is non-fiction, which normally also includes academic prose. However, a mild revival of the use of phrasal verbs with *up* in American non-fiction books has been perceptible since the 1990s. Of the four possible items analyzed in this chapter, i.e. items which can intervene between the verb and the particle in the discontinuous verb-particle combinations, pronouns are by far the most frequent, followed by nouns preceded by an article, usually the definite article, determiners, and lastly, with a very low frequency at all times, nouns preceded by a determiner.

5.5. *Out*

According to *Collins COBUILD Dictionary of Phrasal Verbs* (1989: 449), as well as to Kennedy (1967: 21), the second commonest particle used in verb-particle constructions is *out*. Therefore, in this chapter I shall treat *out* in a similar manner to *up* in the previous section; i.e. I shall first describe its historical and contemporary semantic features, then I will go on with the related corpus findings, in an attempt to check if the uses and diachronic evolution of the VPCs with *out* are consistent with and can corroborate the development of those with *up*.

5.5.1. The Semantics of *out*

Collins COBUILD Dictionary of Phrasal Verbs gives the following definition of *out* as a particle in verb-particle constructions:

> The basic meaning of *out* is to do with movement from the inside of an enclosed space or container to the outside of it. (*Collins COBUILD Dictionary of Phrasas Verbs* 1989: 477)

Although an adverb, in some varieties of English, such as American English or certain non-standard British English dialects, *out* can be used as a preposition with verbs of movement, as in (122) below (cf. *Collins COBUILD Dictionary of Phrasas Verbs* 1989: 477):

(122) He went out the room. (American English)

(123) He went out of the room. (Standard British English)

5.5.1.1. Historical values of *out*

Hiltunen (1983: 212) points out that *ut* (the Old English word for present-day *out*) was frequently used with verbs prefixed by *a-* in the Old English period, e.g. *ut adrifan, ut aflyman, ut agan* 'find out', *ut amæran* 'exterminate', *ut aspeowan, ut aspyrian* 'search, trace'. The reason for this co-occurrence is, according to Hiltunen, the weakening of the prefix *a-*, and as a consequence, its inability to convey the resultative meaning, which was undertaken by the particle *out* instead, as seen in some of the examples above.

A relatively common combination in both Old and early Middle English periods was *berstan ut* 'burst out', as illustrated in the example below (cf. Hiltunen 1983: 212):

(124) Leiden honden on hire þeo þe ihaten weren, & bunden hire þet tet blod *bearst ut* et te neiles.

'Laid hands on those who were called, and tied them up until blood came out of their nails' (*Katherine Group*).

Another combination, with a fairly wide range of meanings in Old and early Middle English was *cuman ut* (cf. Hiltunen 1983: 213):

(125) Þonne cymeð Ilfing eastan in Estmere of ðæm mere Truso standeð in staðe, *cumað ut* samod in Estmere.

'Then the Ilfing comes eastwards into the East Lake, on whose banks Truso stands … come out together into the East Lake' ('The Voyage of Wulfstan' from the Old English *Orosius*);

(126) Daða þa scypmen *common ut* on sæ, þa …

'When the shipmen came out into the sea, then…' (Ælfric's *Homilies*);

(127) þet false gold vorwurðeð þer inne & tet gode gold. *Kumeð ut* brihtre.

'The fake gold perishes in there, and the good gold comes out brighter' (*Ancrene Riwle/Wisse*).

In (125) above, the meaning of the combination is purely literal, namely 'emerge', while in (126) a slight extension can be noticed, the meaning of the construction in this second sentence being 'put out to sea' (cf. Hiltunen 1983: 213). In (127) *ut* acquires even a more abstract sense.

Don ut and *forlætan ut* had the meanings ‘pull out’ and ‘expell’ (cf. Hiltunen 1983: 213). The combination in (128) below also signifies ‘put out to sea’, but in this case the meaning is more metaphorical than in (126) (cf. Hiltunen 1983: 213):

(128) Ond sona ærest, þæs þe heo in scip eodon 7 *ut leton*, þætte astigon wiðorwearde windas.

‘And as soon as they boarded ships and were put out to sea, then contrary winds arose’ (the Old English Bede);

(129) *lette ut* his heorte & forȝet him suluen.

‘let his heart wander and forgot himself’ (*Ancrene Riwle/Wisse*).

In (129) the meaning of the construction *leten ut* is ‘let (something) slip out (of one’s mind or thoughts)’, and this is the earliest attested instance of this combination with this meaning (cf. Hiltunen 1983: 213).

Like *up*, *out*, too, is found in a wider range of verb-particle combinations in early Middle English, as seen in the examples below found by Hiltunen in his corpus (cf. Hiltunen 1983: 213):

(130) betere is þe ðet troddeð wel & *ofsecheð* wel *ut* his owune feblesce

‘better is he who traces and seeks out well his own weakness’ (*Ancrene Riwle/Wisse*);

(131) þer efter *sech* al *ut* & to trodde þine sunnen

‘thereafter seek out all your sins and track them down’ (*ibid*.);

(132) þe hwule þet tu const *siggen out.* & seie al unasked.

‘the while that thou canst confess/speak openly and say all unasked’ (*ibid*.);

(133) te æorl *stæl ut* 7 ferde efter Rodbert eorl of Gloucestre.

‘the earl stole away and went after Robert, Earl of Gloucester’ (*Peterborough Chronicle*).

Hiltunen points out that the meaning of *(of)sechen* is ‘seek (out)’, but, as explained above, in (130) the prefix *of* is insufficient to express the resultative meaning, and therefore, *ut* was used in addition (cf. Hiltunen 1983: 214). The meaning of *siggen ut* in (132) is close to that of *siggen forð*, that is, ‘confess, say openly’, while in (133), the meaning of *stelan ut* is ‘steal

away (secretely)' (cf. Hiltunen 1983: 214). Hiltunen points out that the constructions with *ut* having a metaphorical meaning were not so frequent as the metaphorical *up*-combinations in Old and early Middle English, and that the resultative meaning of *ut* was not as well developed as that of *up* (cf. Hiltunen 1983: 214).

5.5.1.2. Contemporary values of *out*

Present-day *out* conveys a wide range of meanings in addition to its literal use which expresses a movement **outwards** or **out from** something, as, for instance, in *hand out* or *spread out*, and one of these meanings implies a notion of **removal** or **separation**, as in *back out, buy out, cross out, crowd out, keep out, pick out, sell out* (cf. Kennedy 1967: 21). Kennedy notes that this nuance of *out* is particularly obvious in combinations with verbs whose action relates to household chores, such as *air out, brush out, clean out, rinse out, wash out,* or *wring out.* Another meaning noted by Kennedy is that of **completeness** or **finality,** in such combinations as *feather out, hammer out, hew out, leaf out, map out, measure out, plan out, win out, work out,* and so on (cf. Kennedy 1967: 21). Closely related to this meaning, without necessarily implying completeness, is a notion of **openness** or **publicity,** expressed by such combinations as *blaze out, blossom out, boom out, break out* (with the meaning 'to become prevailing'), *call out* ('to call loudly'), *hatch out, pay out* (money), *ring out, weight out,* etc. (cf. Kennedy 1967: 22). Combinations such as *blow out* (meaning 'to extinguish'), *close out, die out, fade out, freeze out, go out* (with the meaning 'to become extinguished'), *strike out, tire out, wear out* convey the idea of **exhaustion** or **extinction** (cf. Kennedy 1967: 22). Kennedy notes that in a small group of combinations both the verb and particle *out* are completely **non-transparent**, that is, *out* has no identifiable meaning, and the verb itself has lost its individual sense, e.g. *find out* ('to discover'), *give out* ('to fail'), *look out* ('to beware'), *make out* ('to comprehend'), *try out* ('to test') (cf. Kennedy 1967: 22).

Bolinger (1971: 104) points out that unlike *up*, *out* is syntactically bounded to certain verbs, and cannot freely enter any combination. Consider for instance the following examples taken from Bolinger:

(134) I helped him out.

(135) *I aided him out.

(136) They developed up some nice reproductions.

(137) They worked out some nice reproductions.

(138) *They developed out some nice reproductions.

The same is true of *rot out* vs. **decay out, carve out* vs. **sculpture out, find out* vs. **discover out*, etc. (cf. Bolinger 1971: 105).

Elenbaas (2007: 21) proposes four meanings of *out*: the literal meaning of **outward direction**, as in *spread out,* or *stick out*, the notion of **removal** or **separation**, as in *clean out* or *pick out*, **disappearance**, and **completion**, e.g. *blow out, fade out, work out, flesh out.*

Tyler & Evans (2007) suggest that "the primary meaning associated with *out* designates a spatial relation in which the trajector TR is exterior to a bounded landmark LM" (Tyler & Evans 2007: 200). In addition, *out* also has a functional component of **non-containment** associated with it (cf. Tyler & Evans 2007: 201). Tyler & Evans identify six clusters of senses with which *out* can be associated, and the first of these is what they term **the Location Cluster**. One of the senses within this cluster is **the Non In Situ Sense**, that is, in this sense, *out* is used to suggest the fact that the entity referred to by the verb is not in its default location, as illustrated in the examples below (cf. Tyler & Evans 2007: 202):

(139) Amy is out sick for the day.

(140) The Robinsons ate out last night.

(141) The workers are out on strike.

Tyler & Evans explain that in (139) the inference is that Amy is in some other place (probably at home) than in the place where she normally spends every day, most probably at her workplace. Similarly, the reading of (141) is that the workers are somewhere else (probably in the streets, at the headquarters of their employer, or in front of the labor authority) than at the workplace – the place where they should normally be. In (140) the inference is that the Robinsons ate somewhere else (obviously at a restaurant or pub) than in the place where they usually eat (obviously their home).

The second sense of *out* within the location cluster is **the No More Sense**, conveyed in situations in which the trajector is a consumable entity kept in a container and when there is no more of it available, as the example below best illustrates (cf. Tyler & Evans 2007: 203):

(142) Have we got any milk left? No, we're (all) out.

With other entities than the 'milk' type, the no more sense can only be achieved by means of the complex preposition *out of*, as in the example below (cf. Tyler & Evans 2007: 203):

(143) We're out of luck / business / sync / love.

An interesting fact is that the no more sense comes in contradiction with the proto-scene of *out*. In (144) below, for instance, where *out* has its literal meaning, the inference is that the scissors are somewhat available for use, hence the contradiction 'no more' vs. 'available-for-use' (cf. Tyler & Evans 2007: 203).

(144) Are the scissors in the drawer? No, they're out on the counter.

The third and last sense of *out* within the location cluster, proposed by Tyler & Evans, is the **Completion Sense**, illustrated by the example below:

(145) This jacket needs to dry out before you wear it again.

Tyler & Evans point out that in sentences such as (145) *out* can be paraphrased by the lexeme *completely*, which confirms the completion meaning added by *out*.

The second cluster of senses proposed by Tyler & Evans for *out* is what they call **The Vantage Point Is Interior Cluster**, in which the viewer or experiencer or the vantage point is located inside the bounded landmark LM, while the trajector TR is exterior to the landmark (cf. Tyler & Evans 2007: 204). Within this cluster, the first sense suggested by Tyler & Evans is the **Exclusion Sense**, i.e. an entity is excluded from the interior environment where the experiencer or viewer is located (cf. Tyler & Evans 2007: 205). This sense is illustrated by the two examples below, taken from Tyler & Evans:

(146) They used a special filter to block out the radio waves.

(147) We use mesh screens to keep the insects out.

If in these examples the exclusion sense is perceived as a **desirable** action, in the following one *out* acquires a distinct nuance which is perceived as **non-desirable** (cf. Tyler & Evans 2007: 205):

(148) The homeless shelter locks out anyone who isn't there by 9pm.

Since the inference in (148) is that the homeless who won't make it to the shelter before 9pm will be literally left outside in the cold, Tyler & Evans argue that one of the functions of the exclusion sense is to contrast the positive nature of the interior environment against the less desirable and sometimes even negative nature of the exterior environment (cf. Tyler & Evans 2007: 205). The exclusion sense is sometimes used in contexts which are not necessarily spatial, and in which the situation is perceived as **unfavorable, punitive** or **disadvantageous,** as the examples below taken from Tyler & Evans (2007: 205) show:

(149) They voted out the unpopular member.

(150) Boots were out of fashion but now they are back in.

(151) The runner was out at third base.

In (149) the situation is unfavorable, as an unpopular candidate was voted out. In (150) boots were in the disadvantageous situation of being out of fashion, while in (151) the punitive nuance is evident, as the player was excluded from the game.

The other sense within the Vantage Point Is Interior Cluster is **the Lack of Visibility Sense** (cf. Tyler & Evans 2007: 206). This means that since the viewer is inside the landmark LM and the trajector TM is exterior, it sometimes happens that the latter is not visible to the viewer, as the example below taken from Tyler & Evans illustrates:

(152) The moment her son went out, Katie started wondering what he was doing.

The interpretation of this sentence is that the trajector, *her son*, is located outside the implicit landmark, *the house,* and as a consequence, the viewer *Katie* no longer has visibility over the trajector, *her son.* Tyler & Evans note that the lack of visibility sense of *out* is conventionalized, and this is evidenced by the following example (cf. Tyler & Evans 2007: 206):

(153) He switched the light out.

Since the light is not physically out of a particular landmark, Tyler & Evans argue that the reader understands the meaning of this sentence only because *out* is conventionally associated with a lack of visibility sense.

Opposed to the previous cluster, where the vantage point is interior to the landmark, **the Vantage Point Is Exterior Cluster** proposed by Tyler & Evans implies that the vantage point is located outside the landmark. In this cluster, **the Visibility Sense** implies that both the vantage point and the trajector are located outside the landmark, and therefore, the TM is visible (cf. Tyler & Evans 2007: 206). As with the lack of visibility sense, the visibility sense, too, is conventionalized, and thus, sentences like (154) below, in which the landmark is indistinct, are fully intelligible (cf. Tyler & Evans 2007: 206):

(154) The sun / moon is out.

Tyler & Evans (2007: 207) note the interesting ability of *out* to convey opposite meanings, i.e. visibility vs. lack of visibility, a contrast which is best illustrated in the two examples below:

(155) The light is out. (cannot be seen → lack of visibility sense)

(156) The sun is out. (can be seen → visibility sense)

An additional meaning of *out* within the Vantage Point Is Exterior cluster which in a way is related to the visibility sense, is **the Knowing Sense**, which implies the knowledge that something is the case (cf. Tyler & Evans 2007: 207). This meaning of *out* is evidenced in sentences such as the ones below taken from Tyler & Evans:

(157) The secret is out.

(158) We figured out the problem.

In neither of the sentences above is there something that can be physically seen. Instead, the inference in (157) is that the secret is known, while in (158) we infer that we know what the problem is. Tyler & Evans explain this particular meaning of *out* by the tight correlation between the notions of seeing and knowing, best evidenced in the example below:

(159) I see / know what you mean.

Like the visibility sense, the knowing sense of *out* is also conventionalized.

Tyler & Evans also propose a **Segmentation Cluster** of senses for *out*, which implies that an entity is separated from the landmark LM where it belongs, in order to be distributed. Thus, **the Distribution Sense** within this cluster is illustrated by sentences such as the ones below (cf. Tyler & Evans 2007: 208):

(160) The player dealt out the cards.

(161) The teacher handed out the test papers.

(162) The chef dishes out the food.

The inference in (162), for instance, is that the chef takes the food out of the pot, which is the landmark in this situation, and segments it into portions, in order to distribute it to the guests.

Finally, Tyler & Evans propose a **Reflexive Sense** of *out*, independent from any cluster, where the landmark and the trajector are one and the same entity, which expands beyond its original boundaries, as the examples below illustrate (cf. Tyler & Evans 2007: 208):

(163) The syrup spread out.

(164) The peacock fanned out its tail.

(165) The boy stretched out his hand.

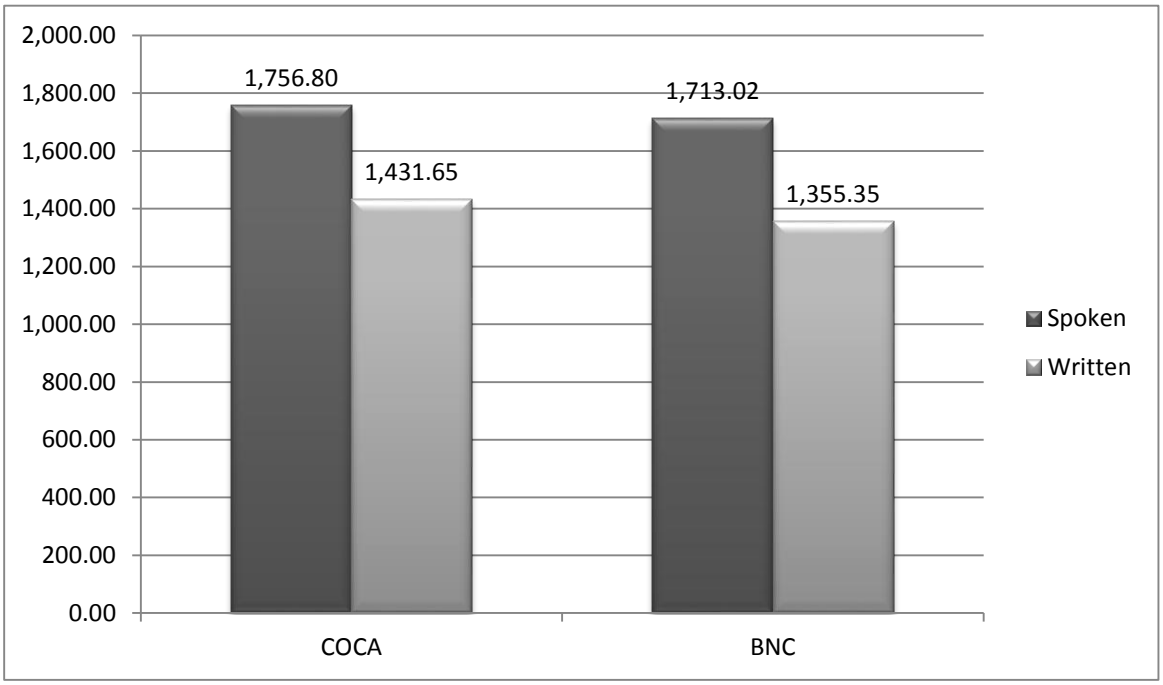
5.5.2. Corpus Findings

Because of the limited length of this thesis, I will consider only the continuous VPCs with *out* in the corpus investigation, and will compare them with the corresponding constructions with *up*, in an attempt to check whether the increasing use of phrasal verbs with *up* is an isolated case, or whether it is only one feature of colloquialization which can be extended over the VPCs with *out*, too, and possibly over phrasal verbs in general.

As the data show, there is an almost identical distribution of the VPCs with *out* in the written and spoken material of the two corpora used, to the corresponding combinations with *up* in the same two corpora. Thus, in both British and American varieties of English the VPCs with *out* are more commonly used in speech, with higher frequencies in the spoken material than in

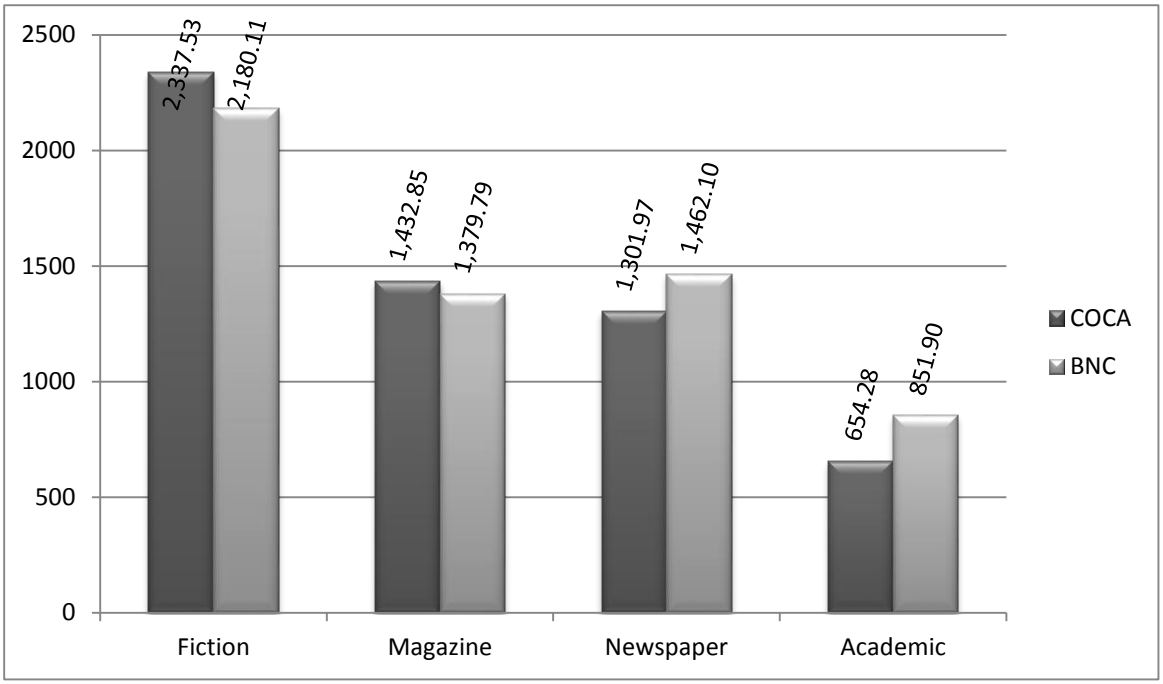
the written sections, as Figure 79 shows. At the same time, the frequencies in American English are slightly higher than the ones in speech and writing found in British English.

Figure 79. Frequencies per million words of the VPCs with *out* in COCA and BNC, in the spoken vs. written material



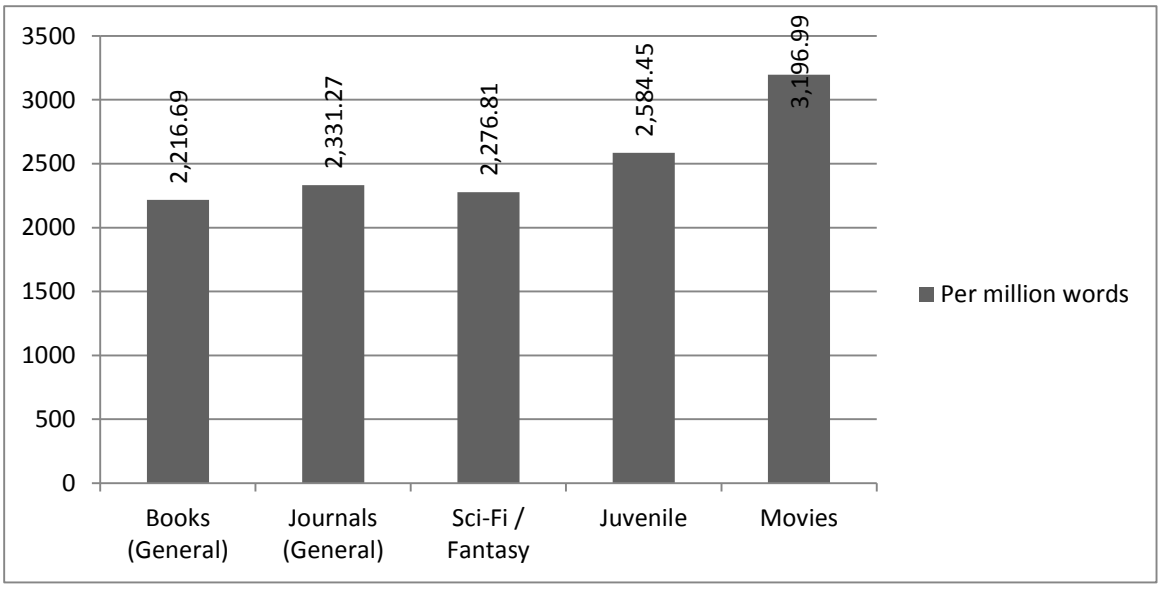
As with *up*, the VPCs with *out* are most widely used in fiction, in both British and American English, as Figure 80 shows. The next frequencies in the American variety of English are, in order, magazine, newspaper and academic prose, but in British English, the frequency of VPCs with *out* in newspapers is higher than the frequency in magazines. While the normalized frequencies in fiction and magazines are higher in American English than in British English, the situation is reversed in the newspaper and academic genres; that is, the frequencies in these two categories are higher in British English than in the American variety. A comparison between Figure 80 and Figure 57 reveals an almost identical pattern in the case of the VPCs with *up*, which might indicate that this distribution among genres is common to phrasal verbs in general, not only to these two particular types of constructions.

Figure 80. Normalized frequencies of the VPCs with *out* in the written material of COCA and BNC, divided by genre



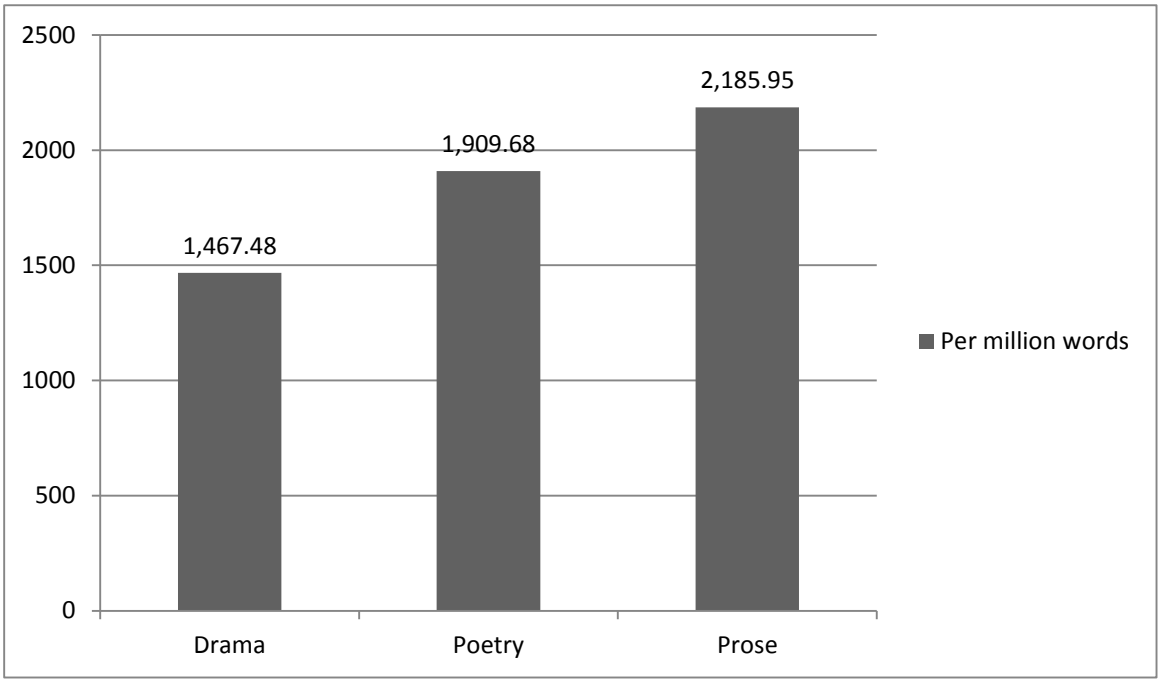
A closer look into the fiction section of COCA discloses the distribution of the continuous VPCs with *out* in the subgenres of this category. Thus, movies lead the way with a stunning frequency per million words of 3,196.99. They are followed in order by youth literature, general journals, science fiction and fantasy literature, and books with a general target audience. This pattern is similar to the usage of the VPCs with *up* in the fiction subgenres of COCA, as a comparison between Figure 81 and Figure 58 shows.

Figure 81. Distribution of the VPCs with *out* in the subgenres of Fiction in COCA



In the fiction subgenres of the BNC, prose has the highest normalized frequency, followed by poetry, and drama (Figure 82).

Figure 82. Distribution of the VPCs with *out* in the subgenres of Fiction in BNC



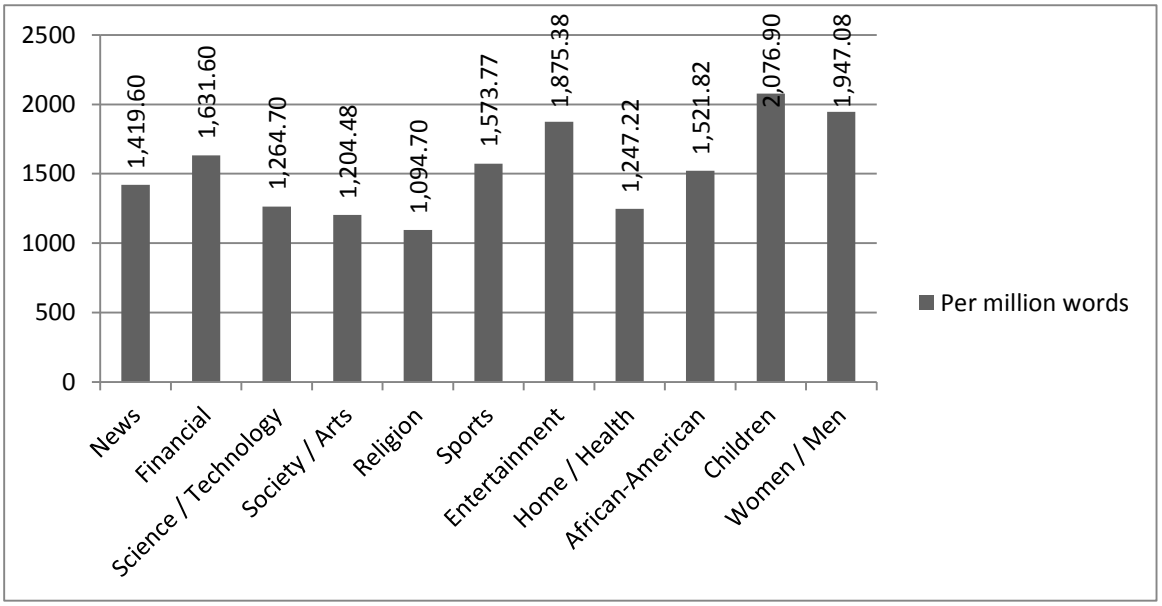
This pattern is no longer similar to that of the combinations with *up* in the same subcategories of the BNC, as *up*-combinations have a higher frequency in drama than in poetry (see Figure 59 for comparison). Below is an example of VPCs with *out* used in poetry, extracted from the fiction section of the BNC:

(166) # It was the masons... of Elm St... # Had nitric acid poured in your pockets? # eyes **gouged out** and maggots in the sockets? # nipples ripped off by a doberman on acid? # intestines **torn out** by a guinea pig that's rabid? # ever been garrotted just for a treat? # It was the masons... of Elm St... # (BNC: CE6)

The highest frequency among the magazine subcategories of COCA is found in children's magazines, followed by men's and women's magazines, entertainment, finance, sports magazines, African American magazines, news magazines, science and technology magazines, home and health magazines, society and arts, and finally, religion magazines (see Figure 83). A similar though not identical usage pattern is found with the continuous VPCs with *up* in the same magazine subcategories of COCA (see Figure 60 for comparison). The only notable difference between the uses of the two constructions is the even higher frequency

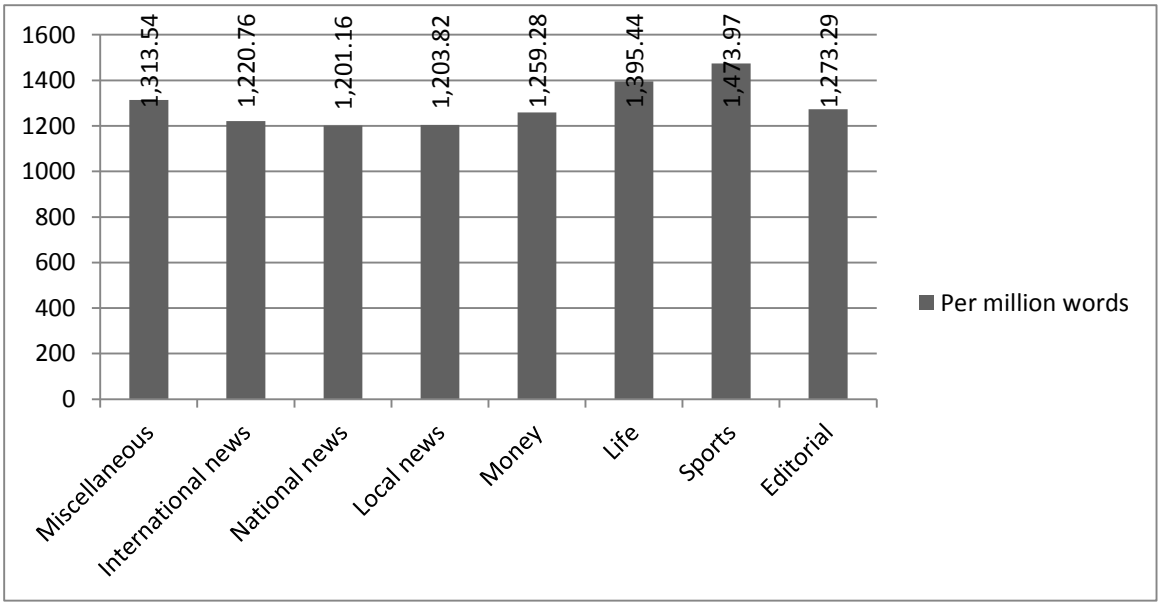
in children's magazines than in men's and women's magazines, in the case of the VPCs with *out*.

Figure 83. Distribution of the VPCs with *out* in the subcategories of Magazine in COCA



In newspapers, the highest frequency is found in sports articles, in American English, followed by lifestyle material, the miscellaneous subgroup, editorials, finance articles, international news material, local news, and finally, national news articles (see Figure 84). The pattern is, again, quite similar to that of the continuous VPCs with *up* in the newspaper subcategories of COCA (Figure 62).

Figure 84. Distribution of the VPCs with *out* in the Newspaper subcategories of COCA

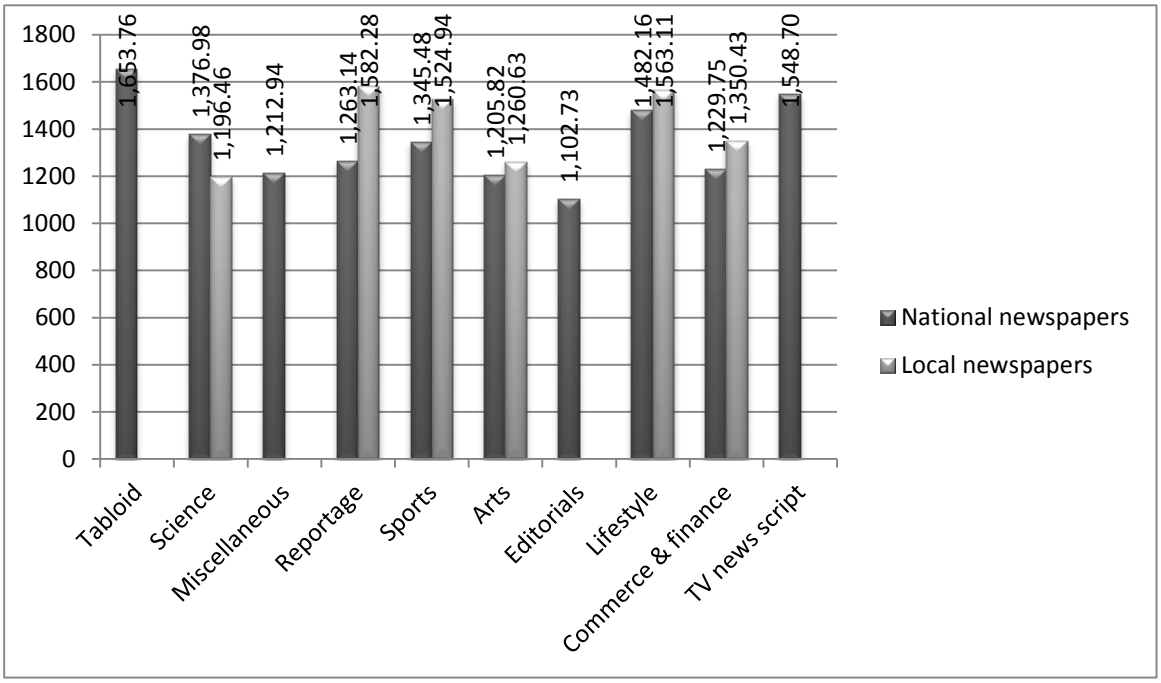


Below is an example of a VPC with *out* employed in a *New York Times* editorial, extracted from the newspaper section of COCA:

(167) Ethiopia **kicked out** the Americans, and Moscow and Washington in effect traded clients, and bases. (COCA)

In British English, the highest use rate of continuous VPCs with *out* is in the tabloid press, followed in order by home and foreign news reportage published in regional and local newspapers; articles on lifestyle, leisure, belief and thought in regional and local newspapers; TV autocue data; sports articles in regional and local newspapers; material on lifestyle published in broadsheet national newspapers; science articles in national newspapers; commerce and finance material in regional and local newspapers; sports articles in national press; reportage published in the national newspapers; articles related to arts published in the regional and local newspapers; commerce and finance material in the national press; the miscellaneous category; arts-related material in the national newspapers; science articles in the regional and local newspapers; and finally, personal and institutional editorials, and letters-to-the-editor published in the broadsheet national newspapers (see Figure 85).

Figure 85. Distribution of the VPCs with *out* in the Newspaper subcategories of BNC

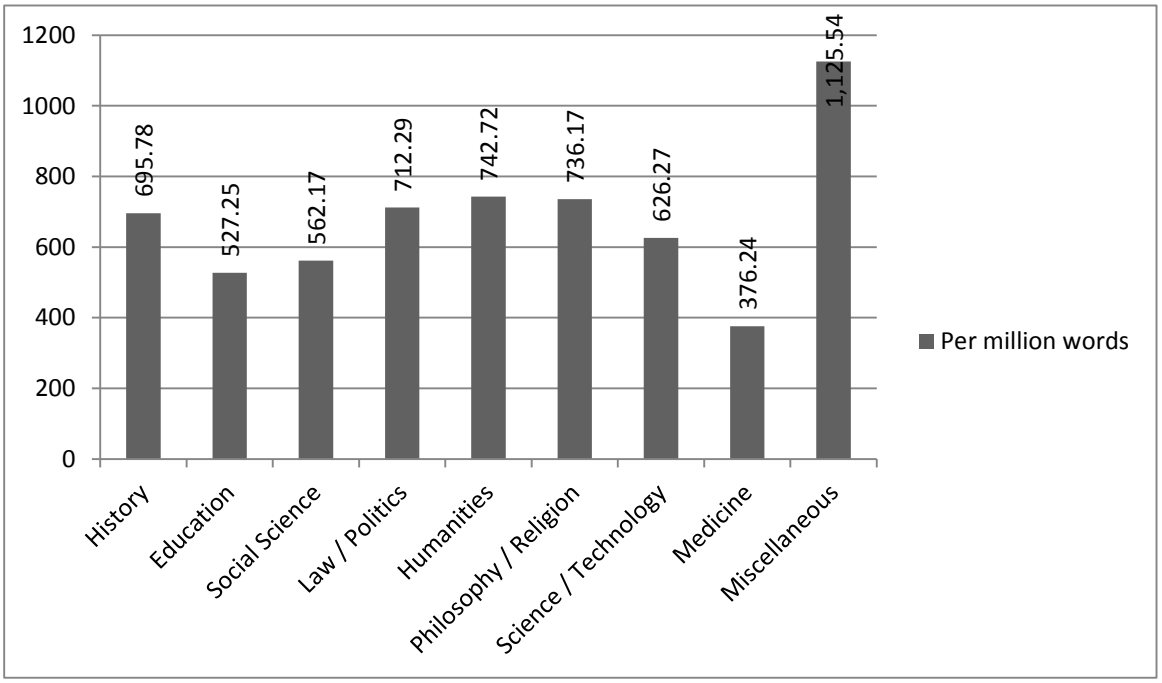


Below is an excerpt from an article published in the British tabloid *The Daily Mirror*, in which continuous verb-particle constructions with *out* are used. The fragment was extracted from the tabloid subcategory of the newspaper section of the BNC.

(168) Antonia said Mellor told her he would like to help her career. He said he would ask Sir David Attenborough to help her, which he did by **checking out** her agent and saying he would **look out** for a film part for her. (BNC: CH1)

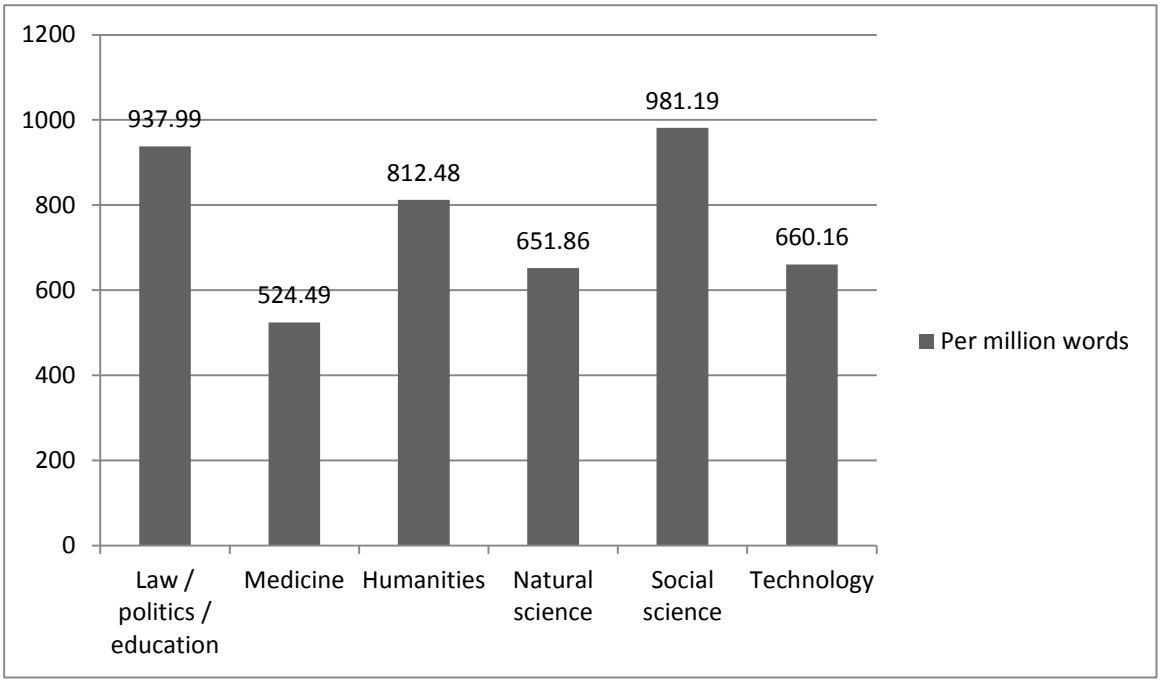
The highest normalized frequency of VPCs with *out* among the subgenres of the academic section of COCA is in the miscellaneous group, followed by the humanities, philosophy and religion, law and politics, history, science and technology, social science, education, and medicine (Figure 86). The distribution is similar to that of the continuous VPCs with *up*, with the only notable difference that philosophy and religion have a somewhat significantly higher frequency in the case of the VPCs with *out* (Figure 64).

Figure 86. Distribution of the VPCs with *out* in the Academic subcategories of COCA



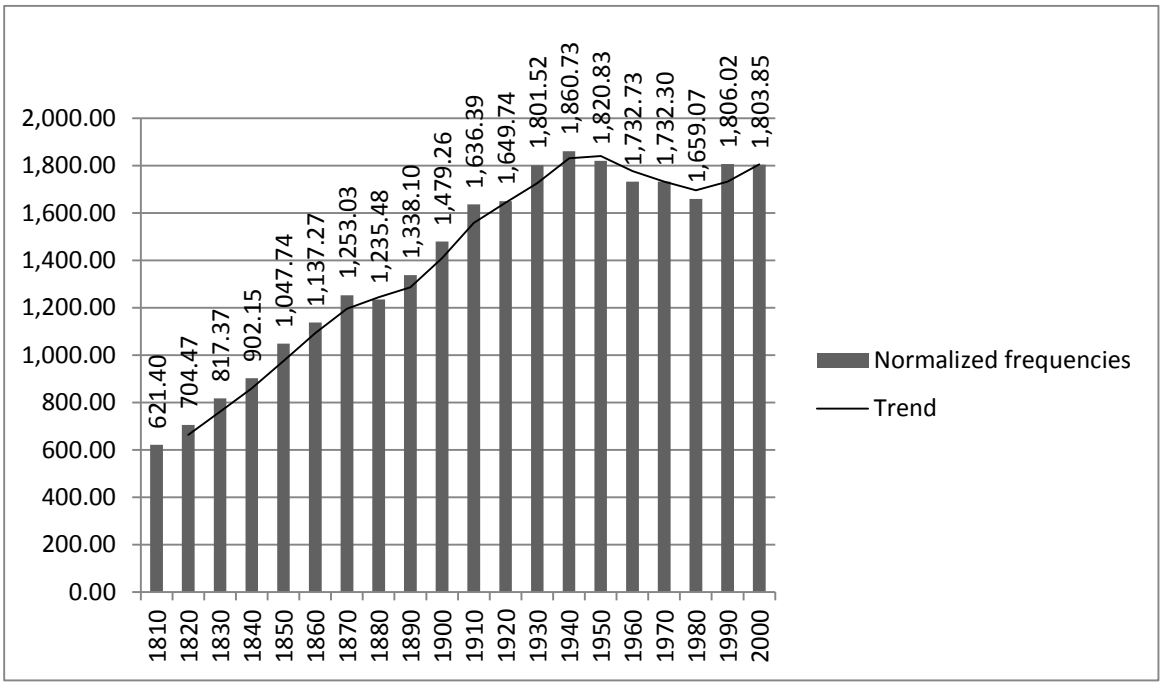
In the BNC, the situation is different, in that social science has the highest normalized frequency in the Academic subcategories, and is followed, in descending order, by the category of law, politics and education, the humanities, technology, natural science, and medicine (Figure 87).

Figure 87. Distribution of the VPCs with *out* in the Academic subcategories of BNC



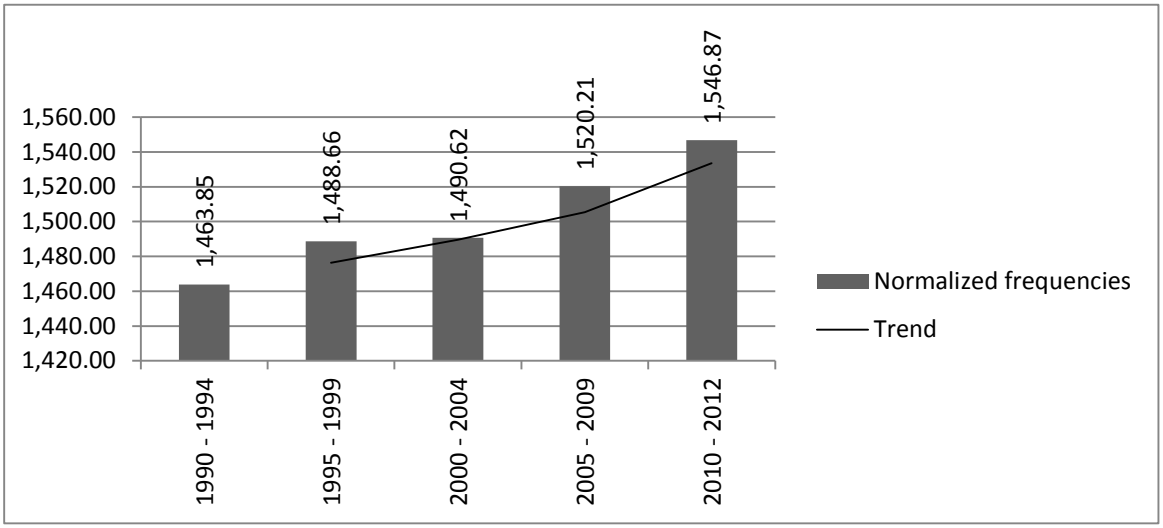
Despite a declining period between the 1940s and the 1980s, the general trend in the use of VPCs with *out* in American English has been clearly ascending over the last two centuries, as Figure 88 suggests.

Figure 88. Diachronic view of the VPCs with *out* in COHA



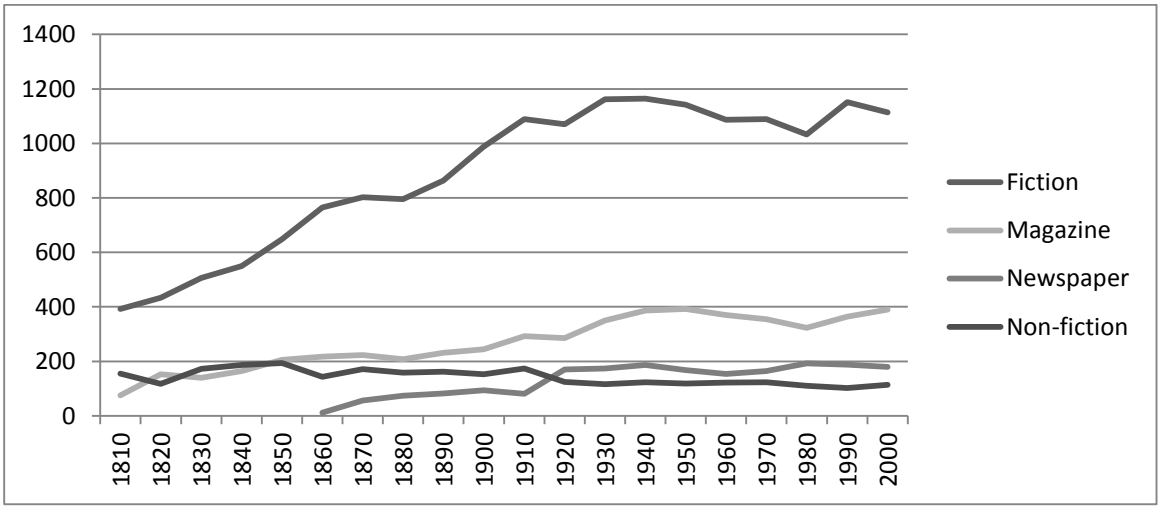
The growing tendency is observable even over the last two decades, a period usually too short to lead to any clear conclusions, but in this case, like the VPCs with *up*, too, the trend is, as Figure 89 shows, clearly increasing.

Figure 89. Diachronic view of the VPCs with *out* in COCA (1990 – 2012)



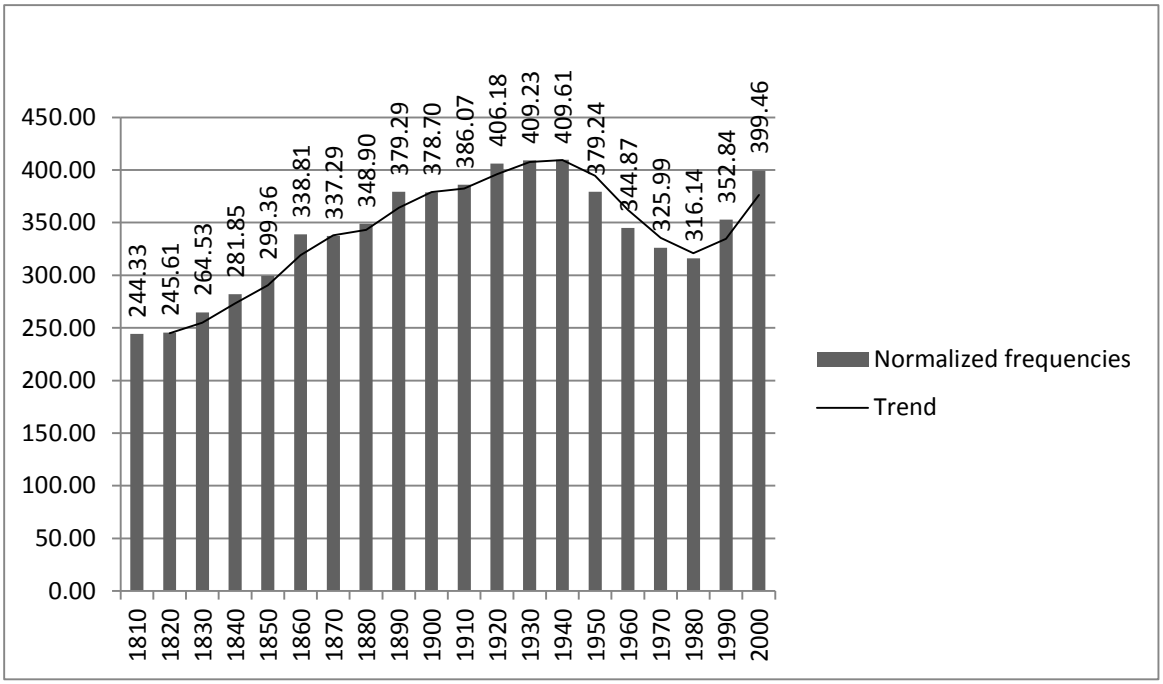
As the genre most open to the informal style, fiction has exhibited the most marked increase in the use of VPCs with *out* in American English over the last two hundred years, followed by magazines and newspapers (Figure 90). As with the VPCs with *up*, the only genre which has seen a decline in the use of the VPCs with *out* is non-fiction, but even so, the latest trend yet seems to have been slightly rising. A comparison of Figure 88, Figure 90 and Figure 89 with Figure 65, Figure 68 and Figure 66, respectively, shows the very similar developments of the combinations with both particles, *up* and *out*.

Figure 90. Evolution of the VPCs with *out* in COHA by genre



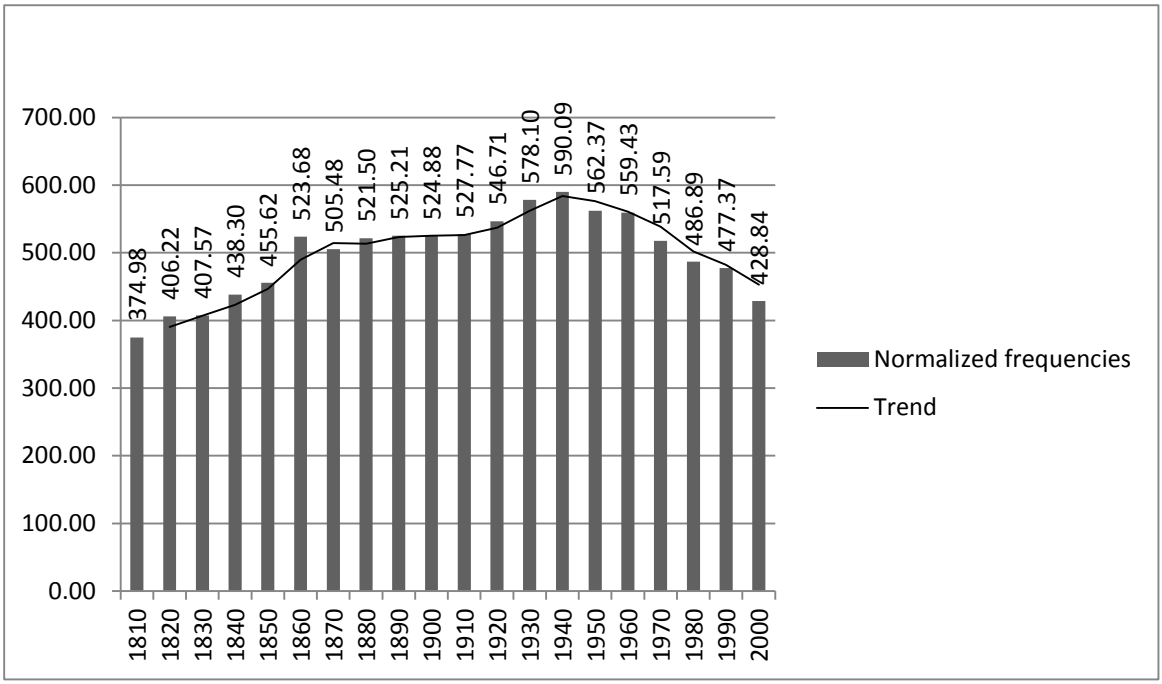
The growing trend in the use of phrasal verbs with *out* observed in COHA is corroborated by the similar results obtained from the Google Books American English corpus, even if the evolution here looks more sinuous, with a more marked decline in the same period of the 1940s – 1980s, but also with a more obvious revival since the 1980s onwards (Figure 91).

Figure 91. Diachronic view of the VPCs with *out* in Google Books American English



However, even if the curves are slightly sharper, overall, the trend line pattern is basically the same, and if we compare the normalized frequency of the 1810s to that of the 2000s, the growth is significant, almost double. Unlike American English, the results from the Google Books British English corpus suggest a somewhat different evolution of the phrasal verbs with *out* in British English, with a growth since the 1810s up to the 1940s, followed by a decline since that decade up to the present (Figure 92). In fact, the development of the combination in the two varieties of English is very similar up to a certain point, in that in both, the combination has seen an increasing tendency since the 1810s up to the 1940s, followed by a decline after that decade. The only significant difference is that in British English the decline seems to have continued up to the present, while in American English, a marked revival of the combination can be noticed since the 1980s. This evolution of the combination with *out* in British English is very similar to that of the combination with *up,* as a comparison between Figure 92 and Figure 69 reveals.

Figure 92. Diachronic view of the VPCs with *out* in Google Books British English



Before moving on to the summary and final conclusions of the thesis, a few concluding remarks on the *out*-combinations are appropriate. The VPCs with *out* show roughly the same synchronic distribution and diachronic evolution as the phrasal verbs with *up*, which might indicate that their increasing use in American English is not an isolated case, but that their evolution is only one part of a more general growth in the use of phrasal verbs. However, *up* and *out* remain the most productive particles in verb-particle constructions, and some relatively recent combinations, such as *freak out* and *screw up* (with its informal meaning 'to bungle'), have had, since their first attestations five or six decades ago, a much more marked boost than the general trend, as Figure 93 and Figure 94 show. The American popularity of phrasal verbs does not seem to have been shared by British English, at least not in the last seven decades, and probably the simple, Latinate equivalents of verb-particle constructions have been preferred in this variety after the 1940s.

Figure 93. ***Freak out*** **in COHA and Google Books Corpora (British and American English; normalized frequencies)**

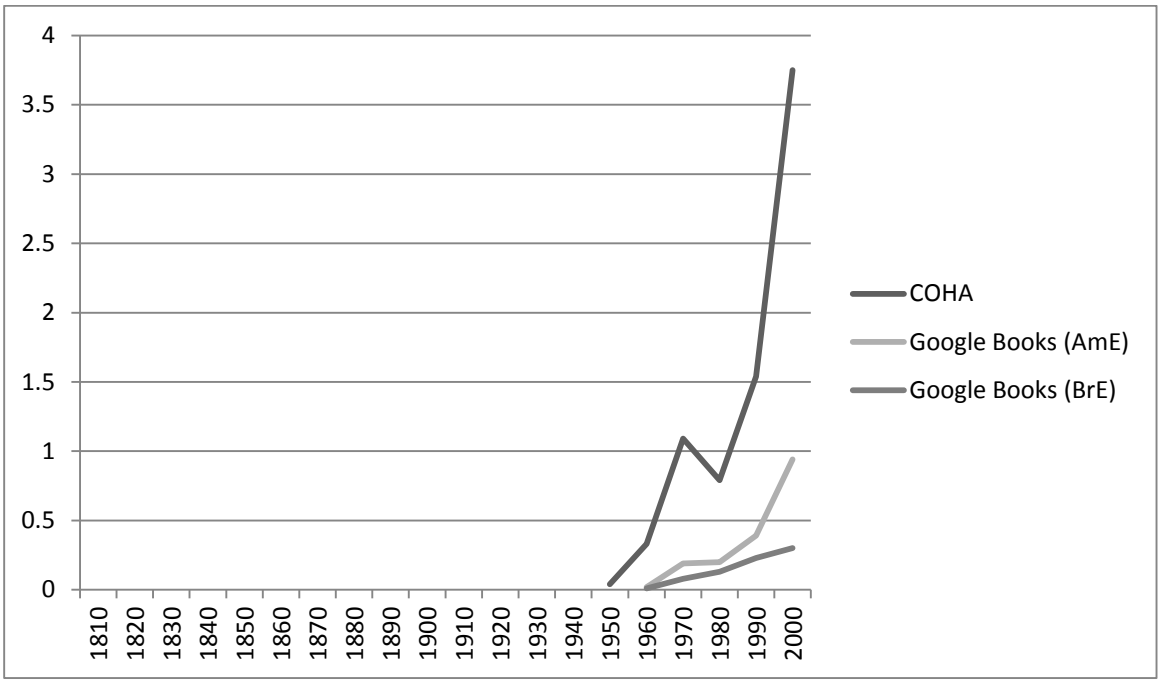
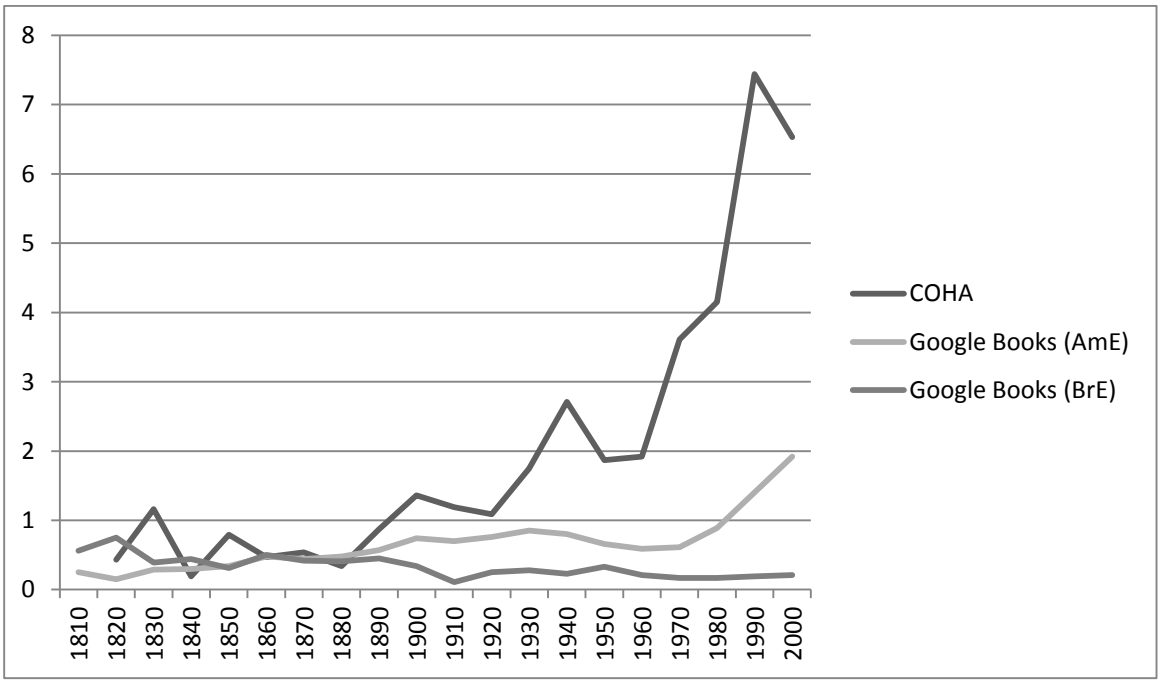


Figure 94. ***Screw up*** **in COHA and Google Books Corpora (British and American English; normalized frequencies)**



6. Conclusions

This thesis dealt with the hypothesis of the colloquialization of written English as put forward by linguists such as Mair (1997, 1998, 1999, 2006, 2007), Hundt (1999, 2009), Smith (2002, 2009), Leech (2004, 2009), Smitterberg (2012) and Westin (2002). An important remark is that in the view of this thesis, as well as on the basis of previous literature, colloquialization is seen not only as a tendency from formality to informality in writing under the influence of speech, but also as writing containing more direct speech. In fact, it is the direct speech fragments in writing where the bulk of spoken features are to be found. A second important remark is that colloquialization should not to be regarded as a single, isolated linguistic phenomenon, but rather as only one part of a much larger socio-cultural phenomenon, involving extra-linguistic factors ranging from politics to economics. The thesis has focused on three assumed colloquialization features, namely the increasing use of contractions and phrasal verbs in writing, and the colloquial uses of *like*, particularly quotative *like*. The corpus findings corroborate the previous research and confirm the colloquialization hypothesis, in that the three features analyzed were found to be spreading in written English. Fiction was found to be the most open genre to colloquialization, as the corpus data shows that the increasing use of the three features is most obvious in this genre, while at the opposite pole is academic prose, which remains the most conservative genre, governed by rigid, highly formal conventions. This corroborates Mair's ranking of academic prose as the prototypical 'slow' genre, i.e. the most resistant to change. Of the two varieties of English considered for investigation, American English was found to be by far the most open to embracing speech features in writing, with marked increasing trends of all three linguistic phenomena analyzed. British English, on the other hand, although not entirely alien to colloquialization, has displayed a debatable tendency from formality to informality in writing. If quotative *like* and contractions such as *lotta* have seen the same marked increase in British English as in American English, the use of phrasal verbs with *up* and *out* has seen a decline after the 1940s, after a growing trend up to that decade. Therefore, phrasal verbs cannot be used as evidence of colloquialization in the case of British English.

Further research is needed for a more comprehensive view of the colloquialization phenomenon. The limited length of this thesis did not allow me to investigate the development of contractions such as *outta* for *out of, let's* for *let us*, *kinda* for *kind of*, *sorta* for *sort of*, and *hafta* for *have to.* Also, other features which might support the

colloquialization hypothesis, and which have not been included in this thesis for the same reasons of length limitation, such as the demise of the archaic preposition *upon* and of the *shall*-future, the decrease in the use of passives and nominalizations in academic prose, and the increasing use of first and second person reference, present tense verbs and the progressive in writing, regardless of genre, still await further and closer analysis. An in-depth study of the language used in new media, such as emails, text messages, blogs, message boards, discussion groups, and online social networks may also provide good insight into recent language developments. Last but not least, a comprehensive interdisciplinary study is required to determine and explain the interplay between colloquialization and the socio-cultural context in which these linguistic changes take place.

Appendix

Table 11. The search strings used in chapter 3

Orthographic form	Search string	Orthographic form	Search string
cannot	can not	*'d*	'd
can't	ca n't	*ain't*	ai n't
n't	n't	*dunno*	du n no
'll	'll	*gimme*	gim me
I'm	I 'm	*gonna*	gon na
it's	it 's	*gotta*	got ta
're	're	*wanna*	wan na

Table 12. The search strings used in chapter 4

Form	Function	Search string	Example	Meaning / Use
(CONJ) + PRON + *be* + *like*	Quotative	[c*] [p*] [be] like ,|'	*(And) I was like, "wow".*	Introduces direct speech or non-verbal communication
Be + *like* + TO *infinitive*	Adverb	I|you|he|she|we|they [be] like to [v?i*]	*She was like to faint.*	be likely to, be ready to, be on the verge of
…, *like, …* / …, *like.*	Filler	, like , / , like .	*I, like, didn't say anything.*	Intentionally inserts a break
	Subordinating conjunction	like.[cs*]	*You look like you were dead.*	as though, as if, as

Table 13. The search strings used in chapter 5

Syntactic form	Query syntax	Example
V + *up* (+ DO)	[vv*] up	*Pick up the pack*
V + ART + N + *up*	[vv*] [at*] [nn*] up	*Turn the heat up*
V + DET + N + *up*	[vv*] [dd*] [nn*] up	*Clean this mess up*
V + DET + *up*	[vv*] [dd*] up	*Clean this up*
V + PRON + *up*	[vv*] [pp*] up	*Pick it up*
V + *out* (+ DO)	[vv*] out	*She freaked out*

Bibliography

Beal, J. C. (2004). *English in Modern Times 1700-1945.* London: Arnold.

Biber, D. and Conrad, S. (2009). *Register, Genre, and Style.* Cambridge: Cambridge University Press.

Biber, D. and Finegan, E. (1989). Drift and the Evolution of English Style: A History of Three Genres. *Language, 65*(3), 487-517.

Bolinger, D. (1971). *The Phrasal Verb in English.* Cambridge: Harvard University Press.

Buchstaller, I. and Van Alphen, I. (2012). Preface: Introductory remarks on new and old quotatives. In I. Buchstaller and I. Van Alphen, *Quotatives: Cross-linguistic and cross-disciplinary perspectives* (pp. XI-XXX). Amsterdam: John Benjamins.

Burgess, A. (1986). *A Clockwork Orange.* Harmondsworth: Penguin Books.

Collins COBUILD Dictionary of Phrasal Verbs. (1989). London: Collins Publishers.

Collins English Dictionary Online. < http://www.collinsdictionary.com/ >. Accessed February 2013.

Dalzell, T. and Victor, T. (2008). *The Concise New Partridge Dictionary of Slang and Unconventional English.* New York: Routledge.

Davies, M. *BYU-BNC: British National Corpus*. < http://corpus.byu.edu/coca/ >. Accessed September 2012.

Davies, M. *Corpus of Historical American English.* < http://corpus.byu.edu/coha/ >. Accessed July 2012.

Davies, M. *Corpus of Contemporary American English*. < http://corpus.byu.edu/coca/ >. Accessed July 2012.

Davies, M. *Google Books: American English Corpus*. < http://googlebooks.byu.edu/ >. Accessed January 2013.

Davies, M. *Google Books: British English Corpus*. < http://googlebooks.byu.edu/ >. Accessed January 2013.

Dehé, N. (2002). *Particle Verbs in English: Syntax, information structure and intonation.* Amsterdam: John Benjamins.

Elenbaas, M. (2007). *The Synchronic and Diachronic Syntax of the English Verb-Particle Combination.* Utrecht: LOT.

Faulkner, W. (1942). The Bear. In W. Faulkner, *Go Down, Moses and other Stories* (pp. 135-236). London: Chatto & Windus.

Fowler, R. (1991). *Language in the News: Discourse and Ideology in the Press.* London: Routledge.

Gries, S. T. (2003). *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement.* New York: Continuum.

Hiltunen, R. (1983). *The Decline of the Prefixes and the Beginnings of the English Phrasal Verbs: The Evidence from some Old and Early Middle English Texts.* Turku: Turun Yliopisto.

Hundt, M. and Mair, C. (1999). "Agile" and "Uptight" Genres: The Corpus-based Approach to Language Change in Progress. *International Journal of Corpus Linguistics, 4*(2), 221-242.

Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. In N. Dehé, & e. al., *Verb-Particle Explorations* (pp. 67-94). Berlin: Mouton de Gruyter.

James, H. (2008). *Roderick Hudson.* Stilwell: Digireads.com Publishing.

Kennedy, A. G. (1967). *The Modern English Verb-Adverb Combination.* New York: AMS Press.

Leech, G. (2004). *Meaning and the English Verb.* London: Pearson Education Limited.

Leech, G., Hundt, M., Mair, C. and Smith, N. (2009). *Change in Contemporary English: A Grammatical Study.* Cambridge: Cambridge University Press.
Mair, C. (1997). The Spread of the Going-to-Future in Written English: A Corpus-Based Investigation into Language Change in Progress. (R. Hickey, & S. Puppel, Eds.) *Language History and Linguistic Modelling. A Festschrift for Jacek Fisiak on His 60th Birthday*, 1537-1543.
Mair, C. (1998). Corpora and the study of the major varieties of English: Issues and results. In H. Lindquist, S. Klintborg, M. Levin, & M. Estling, *The Major Varieties of English: Papers from MAVEN 97, Växjö 20-22 November 1997* (pp. 139-157). Växjö: Växjö University.
Mair, C. (2006). *Twentieth-Century English: History, Variation and Standardization.* Cambridge: Cambridge University Press.
Mair, C. (2007). Change and variation in present-day English: integrating the analysis of closed corpora and web-based monitoring. In M. Hundt, & e. al., *Corpus Linguistics and the Web* (pp. 233-247). Amsterdam: Editions Rodopi.
McEnery, T. and Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice.* Cambridge: Cambridge University Press.
Richards, J. C. and Schmidt, R. (2010). *Longman Dictionary of Language Teaching & Applied Linguistics. Fourth Edition.* Harlow: Pearson Education.
Schourup, L. C. (1985). *Common Discourse Particles in English Conversation.* New York: Garland Publishing.
Schourup, L. and Butters, R. (1982). Quoting with Go 'Say'. *American Speech, 57*(2), (pp. 148-149).
Smith, N. (2002). Ever Moving on? The Progressive in Recent British English. (A. Smith, P. Peters, and P. Collins, Eds.) *New Frontiers of Corpus Research*, 317-330.
Smitterberg, E. (2012). Colloquialization and NOT-contraction in nineteenth-century English. In M. Markus, & al., *Middle and Modern English Corpus Linguistics: A multi-dimensional approach* (pp. 191-206). Amsterdam: John Benjamins.
Stenström, A.-B., Andersen, G. and Hasund, I. K. (2002). *Trends in Teenage Talk: Corpus compilation, analysis and findings.* Amsterdam: John Benjamins.
Stevenson, R. L. (1957). *Kidnapped.* London: Oxford University Press.
Thim, S. (2012). *Phrasal Verbs: The English Verb-Particle Construction and its History.* Berlin: De Gruyter Mouton.
Tottie, G. (2002). *An Introduction to American English.* Oxford: Blackwell Publishing.
Tyler, A. and Evans, V. (2007). *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition.* Cambridge: Cambridge University Press.
Vandelanotte, L. (2012). Quotative 'go' and 'be like': Grammar and grammaticalization. In I. Buchstaller, & I. Van Alphen, *Quotatives: Cross-linguistic and cross-disciplinary perspectives* (pp. 173-202). Amsterdam: John Benjamins.
Wentworth, H. and Flexner, S. B. (1975). *Dictionary of American Slang. Second Supplemented Edition.* New York: Thomas Y. Crowell, Publishers.
Westin, I. (2002). *Language Change in English Newspaper Editorials.* Amsterdam: Rodopi.